

STRUCTURAL STUDIES OF: PAM (PHENYLALANINE AMINOMUTASE),
AND BADA (BENZOATE COENZYME A LIGASE); PURIFICATION AND
CRYSTALLIZATION TRIALS OF: mSNAPc (SMALL NUCLEAR RNA ACTIVATING
PROTEIN), AND TF8 (TFIIIB BR1-TBP TRIPLE FUSION)

By

Susan Marie Strom

A DISSERTATION

Submitted to
Michigan State University
In partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Chemistry

2012

ABSTRACT

STRUCTURAL STUDIES OF: PAM (PHENYLALANINE AMINOMUTASE), AND BADA (BENZOATE COENZYME A LIGASE); PURIFICATION AND CRYSTALLIZATION TRIALS OF: mSNAPc (SMALL NUCLEAR RNA ACTIVATING PROTEIN), AND TF8 (TFIIIB BRF1-TBP TRIPLE FUSION)

By

Susan Marie Strom

Pantoea agglomerans phenylalanine aminomutase (*PaPAM*) is an enzyme that reacts with (2R)- α -phenylalanine to produce (3S)- β -phenylalanine in the biosynthetic production of the antibiotic Andrimid. The mechanism by which this class of enzymes achieves this transformation is debated. The crystal structure of the (3S)- β -phenylalanine bound PaPAM was determined with both (2R)- α -phenylalanine and (3S)- β -phenylalanine bound to the active site providing evidence that this class of enzymes utilizes an amino-group alkylation pathway.

Benzoate Coenzyme A (CoA) Ligase from *Rhodospseudomonas palustris* (*BadA*) catalyzes the ligation of Coenzyme A to a variety of benzoic acids in the presence of adenosine triphosphate (ATP). Benzyl-CoAs are useful in the biosynthesis of small molecules. The crystal structure was therefore determined with various natural and unnatural substrates bound to the active site and within water exposed channels. These structures demonstrate the mode by which this enzyme catalyses the reaction; aiding in our understanding of its mode of action and our ability to increase the promiscuity of the enzyme to produce benzyl-CoA derivatives.

Small Nuclear RNA Activating Protein (SNAPc) is a human nuclear transcription factor composed of five subunits that activates the transcription of small nuclear RNA by recruiting RNA polymerase I, II or III to various promoters. Although different polymerases are recruited, it is the presence of the same element, the proximal sequence element (PSE), upstream of the transcription start site that recruits SNAPc. The U1 promoter contains such an element to which

SNAPc recruits RNA polymerase II (Pol II). In the case of the U6 promoter, the presence of a TATA box additionally recruits the TATA Binding Protein (TBP) and ultimately RNA Polymerase III (Pol III) is activated. However there is no evidence that it is the presence of TBP that would discriminate between Polymerase II and Polymerase III recruitment. To understand this phenomenon, a truncated version of SNAPc was co-expressed in *E. coli* for crystallization studies. Though a complex of the four proteins could be produced, attempts to crystallize it were unsuccessful.

Transcription factor IIIB (TFIIIB) recruits Pol III in budding yeast such as *Saccharomyces cerevisiae* (*Sc*). In gene internal promoters containing an A Box element, Transcription Factor IIIA (TFIIIA) is first recruited which in turn recruits TFIIIC which then recruits TFIIIB. The gene internal promoters such as those for the tRNA containing an A Box element recruit TFIIIC directly, which in turn recruits TFIIIB. In the case of the U6 snRNA promoter which contains a TATA Box element it was shown that TFIIIB is recruited directly to the promoter via the subunit TBP. In order to facilitate crystallization, a BRF1-TBP triple fusion was generated containing the amino and carboxyl termini of Brf1 with TBP inserted in between. This triple fusion was shown to have the same activity as the separate units; however it proved to be unsuitable for crystallization.

Copyright by
SUSAN MARIE STROM
2012

I dedicate this work to my wonderful and loving husband Kevin.

ACKNOWLEDGEMENTS

Dr. James H. Geiger, thank you for being my advisor. I appreciate the guidance and space you gave me to explore this strange microscopic world. Dr. Bill Henry, thank you for being my second reader and for all of your help over the years. Thank you also to my committee members: Dr. Kevin Walker, Dr. Babak Borhan and Dr. David Weliky. Dr. Preeti Dhar, my undergraduate advisor and mentor, I thank you for all of your guidance. Dr. Stacy Hovde, I thank you for all of your help over the years.

Thanks to all of my undergraduate assistants, particularly Justin, Eric, Kelsey and Shaun. I'll never forget the Geiger Lab Olympics. For the record, Justin won the gold!

I would also like to thank those who helped me in my work off campus. I thank the Kassavetis lab for all of their support in working with the TFIIB fusions. Without the aid of the Ferguson-Miller lab I never could have completed all those crystallization screens. I thank the Castalino group for their aid in the conantokin and plasminogen binding protein work. I also thank all of the staff at LS-CAT. There were several times you had to be called in the middle of the night to fix the beam and even more occasions where you helped me collect data remotely. It was a pleasure working with all of you.

To my lab mates and friends I also give my thanks: Lei, Xiaoafei, Sara, Fang, Andy, Blanka, Dorothy, Rafida, Remie, Meisam, Uday, Chelsea and Danielle. All of us worked together on a project at one time or another, and sometimes the project actually worked! When you are famous remember that I taught you everything you know.

I thank my family: my husband Kevin, my little one Martin, my "Michigan Mom" Kathy, my Mom, my Dad, all of my sisters and brothers, Tom, Aunt Patti, and everyone else for their encouragement and support. I love all of you and thank God for you!

TABLE OF CONTENTS

LIST OF TABLES.....	ix
LIST OF FIGURES.....	xi
LIST OF ABBREVIATIONS.....	xx
CHAPTER I: <i>Pa</i>PAM	
I.1. Background	
I.1.1. Co-opting biosynthetic pathways	1
I.1.2. Phenylalanine Aminomutases.....	2
I.1.3. Ammonia Lyase Mechanistic Studies.....	5
I.1.4. Structures of MIO-containing enzymes.....	10
I.2. Experimental Procedures	
I.2.1. Crystallization of <i>Pantoea agglomerans</i> Phenylalanine Aminomutase.....	24
I.2.2. Structure Determination	26
I.3. Results and Discussion	
I.3.1. Crystal Structure of <i>Pantoea agglomerans</i> Phenylalanine Aminomutase.....	28
References.....	57
CHAPTER II: <i>Bad</i>A	
II.1. Background	
II.1.1. Coenzyme A Ligases Structure and Function.....	63
II.1.2. Benzoate and Benzoate derivative Coenzyme A Ligase Structures	66
II.2. Experimental Procedures	
II.2.1. Crystallization of <i>Rhodopseudomonas palustris</i> Benzoate-Coenzyme A Ligase.....	75
II.2.2. Soaking and Co-crystallization Experiments.....	77
II.2.3. Structure Determination.....	79
II.3. Results and Discussion	
II.3.1 Crystal Structures of <i>Rhodopseudomonas palustris</i> Benzoate-Coenzyme A Ligase.....	83
References.....	107
CHAPTER III: <i>SNAPc</i>	
III.1. Background	
III.1.1. RNAPII and RNAPIII.....	110
III.1.2. Small Nuclear RNA Promoters.....	111
III.1.3. Small Nuclear RNA Activating Protein Complex (<i>SNAPc</i>).....	112
III.1.4. Previous studies of mini-Small Nuclear RNA Activating Complex (<i>mSNAPc</i>).....	112
III.2. Experimental Procedures	
III.2.1. Co-expression of <i>mSNAPc</i>	114
III.2.2. Initial Standard Purification of <i>mSNAPc</i>	115
III.2.3. Alternative <i>SNAPc</i> Purification Protocol I.....	116

III.2.4. GST Resin Binding Time Optimization.....	116
III.2.5. Optimization of Length of Time of Thrombin Digestion.....	116
III.2.6. Further Optimization of Buffers.....	117
III.2.7. Alternative SNAPc Purification Protocol II.....	118
III.2.8. Crystallization Trials.....	118
III.2.9. His-tagged SNAP50.....	119
III.2.10. Purification of Co-expressed His-tagged SNAP50, GST-tagged SNAP190 (1-505), SNAP43 and SNAP19 (N-His γ 4).....	120
III.2.11. SNAP190 (1-131), (1-135), (1-255), (1-260), (1-265).....	121
III.2.12. Tagless SNAP190 (1-505).....	123
III.2.13. Maltose Binding Protein Tagged SNAP190 (1-505).....	124
III.2.14. Maltose Binding Protein Tagged SNAP190 (1-505) with Thrombin Linker.....	126
III.2.15. Maltose Binding Protein Tagged SNAP190 (1-505) with Smt3 Linker.....	126
III.2.16. Maltose Binding Protein Tagged SNAP190 (1-131) with Smt3 Linker.....	129
III.2.17. Maltose Binding Protein Tagged SNAP50 with Smt3 Linker.....	129
III.2.18. SNAP19 Truncations.....	132
III.2.19. Surface Entropy Reduction Mutations of SNAP190 (1-505).....	133
III.2.20. Maltose Binding Protein tagged SNAP190 (Δ 131-260) with SMT3 cut site and various linkers.....	133
III.2.21. Maltose Binding Protein tagged SNAP190 (260-505) with SMT3 cut site.....	136
III.3. Results and Discussion	
III.3.1. Purification and attempted crystallization of the SNAPc complex.....	137
References.....	157
CHAPTER IV: TFIIB Brf1-TBP Triple Fusions	
IV.1. Background	
IV.1.1. TFIIB.....	160
IV.1.2. Creation of the TFIIB-Brf1-TBP Triple Fusion.....	161
IV.2. Experimental Procedures	
IV.2.1. Growth Optimization.....	164
IV.2.2. Purification Optimization.....	164
IV.3. Results and Discussion	
IV.3.1. Purification and attempted crystallization of TBP-Brf1 fusions.....	167
IV.3.2. Crystallization of TF8-DNA complexes.....	172
References.....	175

LIST OF TABLES

Chapter I

Table I.1	Sequence alignment of class I lyase-like family members using Clustal 2.1 (30). Some key residues have been highlighted. Proteins are identified by their PDB ID codes followed by the abbreviations as follows: <i>Pp</i> HAL: <i>Pseudomonas putida</i> histidine ammonia lyase; <i>Rs</i> TAL: <i>Rhodobacter sphaeroides</i> tyrosine ammonia lyase; <i>Pa</i> PAM: <i>Pantoea agglomerans</i> phenylalanine aminomutase; <i>Sg</i> TAM: <i>SgcC4</i> L-tyrosine 2,3-aminomutase; <i>Pc</i> PAL: <i>Petroselinum crispum</i> phenylalanine ammonia lyase; <i>Tc</i> PAM: <i>Taxus canadensis</i> phenylalanine aminomutase; <i>Rt</i> PAL: <i>Rhodospiridium toruloides</i> phenylalanine ammonia lyase; <i>Av</i> PAL: <i>Anabaena Variabilis</i> ATCC 29413 phenylalanine ammonia lyase; <i>Np</i> PAL: <i>Nostoc punctiforme</i> ATCC 29133 phenylalanine ammonia lyase.....11
Table I.2	Data-Collection and Structure-Refinement Statistics for <i>Pa</i> PAM.....27
Table I.3	Multiple sequence alignment using Clustal 2.1(32) of proteins identified by a BLAST search (55) as having high sequence similarity to <i>Pa</i> PAM suggests four yet to be characterized proteins [<i>Vibrionales bacterium</i> SWAT-3 (V. bact), <i>Bacillus Subtilis</i> (B.subtl), <i>Klebsiella pneumoniae</i> 342 (K.pneu) and <i>Burkholderia rhizoxinica</i> (B.rhiz)] as possible PAMS with stereoselectivity analogous to that of <i>Pa</i> PAM. <i>EncP</i> from <i>Streptomyces maritimus</i> is abbreviated S. marit. Potentially important active site residues are highlighted.....52

Chapter II

Table II.1	Data collection and structure refinement statistics for the structures containing benzoic acid, p-toluic acid and 2-fluorobenzoic acid.....80
Table II.2	Data collection and structure refinement statistics for the structures containing o-toluic acid, 2-furoic acid and thiophenic acid.....81
Table II.3	Data collection and structure refinement statistics for the structure containing benzoic acid ligated to adenosine monophosphate (AMP).....82
Table II.4	Comparison of different conformations found in ATP dependent CoA ligases....86

Table II.5	Multiple sequence alignment of BadA, BCL _M (BCL _m), CBL and ACSM2A using Clustal 2.1 (18). The conserved A8 domain is highlighted in yellow. The conserved A10 domain is highlighted in blue (19). Residues that are part of the substrate binding pocket are marked in light grey. Residues involved in binding the carboxylate are highlighted in dark grey. Residues involved in AMP binding are highlighted in green. Residues involved in the hinge movement between the N- and C-terminals are marked in red. Underlined residues are involved in CoA binding or are suspected to be involved in CoA binding as evident by structural overlays with CBL where structures of a CoA bound enzyme do not exist (BadA and BCL _M).....	87
-------------------	--	----

Chapter III

Table III.1	Buffers used in the optimization of the purification of mSNAPc.....	117
Table III.2	Buffers used in the purification of mSNAPc.....	140

Chapter IV

Table IV.1	DNA strands derived from SymSelex and SymSelex2 sequences. The lowercase “p” represents a site of phosphorylation.....	173
-------------------	--	-----

LIST OF FIGURES

Chapter I

- Figure I.1** Top: The conversion of (2S)- α -phenylalanine to (3R)- β -phenylalanine (purple) by *Taxus canadensis* phenylalanine aminomutase (*TcPAM*) and to (3S)- β -phenylalanine (blue) by *Pantoea agglomerans* phenylalanine aminomutase (*PaPAM*). Bottom: The antibiotic Andrimid and anti-cancer drug Taxol both contain a β -phenylalanine component (highlighted in blue and purple) and contain phenylalanine aminomutases within their biosynthetic pathways (*TcPAM* and *PaPAM*, respectively). For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation...3
- Figure I.2** PAMs are a member of the class I lyase-like family perform similar chemical transformations and include tyrosine aminomutases (TAMs), tyrosine ammonia lyases (TALs), phenylalanine ammonia lyases (PALs), and histidine ammonia lyases (HALs).....6
- Figure I.3** The autocatalytic formation of the MIO in *PaPAM* from the sequence threonine-serine-glycine is adapted from that of HAL (25).....7
- Figure I.4** Top: Structure of the MIO. Bottom: Two proposed mechanisms for the conversion of substrate to product in a generic aminomutase. X represents hydrogen in the case of phenylalanine and a hydroxyl group in the case of tyrosine. For a generic ammonia ligase the reaction would stop with the production of *trans*-cinnamate in the case of PALs or *trans*-coumerate in the case of TALs.....9
- Figure I.5** The formation of urocanate in HALs as presented by Schwede et. al. (20) inspired the Friedel-Crafts type mechanism in PALs and PAMs.....15
- Figure I.6** Ribbon diagrams overlaid of RtPAL (green) with PpHAL (blue) highlight the capping domain present in plant PALs and PAMs.....17
- Figure I.7** Key residues in the active site of RtPAL are shown. Residues from monomers 1, 2 and 3 are shown in yellow, green and blue respectively. Cinnamate bound in the active site is shown in magenta. The MIO with an apparent ammonia adduct bound is highlighted in red.....17
- Figure I.8** Active site of *SgTAM* with bound inhibitor (2S,3S)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid (blue) (left, PDB ID 2RJR) and the substrate mimic (3R)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (red) (right, PDB ID 2RJS) suggests an amino-group alkylation pathway.....20

Figure I.9	The key TcPAM residues that interact with cinnamate. Residues of “subunit A” are colored blue and residues of “subunit B” are colored magenta.....	21
Figure I.10	Crystals of <i>Pantoea agglomerans</i> Phenylalanine Aminomutase.....	25
Figure I.11	<i>Pa</i> PAM monomer (green) showing the inner and outer loop regions (as marked), MIO (blue spheres) and (3S)- β -phenylalanine (red spheres). The active site rests at the monomer’s surface at the end of a long four helix bundle.....	29
Figure I.12	The PaPAM tetramer showing the relationship between the four monomers is that of a head-to-tail arrangement. Four active sites rest at the interfaces created by the tetramer. The residues from three monomers contribute to the active sites including residues from the inner loop of the MIO containing monomer to the outer loop residues of an adjacent monomer. The MIO is highlighted in blue spheres. The (3S)- β -phenylalanine is highlighted in red spheres and (2S)- α -phenylalanine is highlighted in orange spheres which are seen overlapping with those of the (3S)- β -phenylalanine.....	30
Figure I.13	Each of the four MIOs (blue spheres) within the tetrameric form of PaPAM is positioned at the end of a long four-helix bundle (see green helixes on left hand side of image) which are believed to stabilize negative charges generated on the MIO during the amino-group alkylation pathway. Ribbon diagrams of the four separate monomers are shown in magenta, green, yellow and aqua. The (3S)- β -phenylalanine is highlighted in red spheres and (3S)- α -phenylalanine in highlighted in orange spheres.....	31
Figure I.14	The PaPAM tetramer showing the relationship between the four monomers and the interactions between the inner and outer loop regions and the MIO active site. The inner loop of the monomer containing the MIO is packed between the active site and the outer loop region of a neighboring monomer. The MIO is highlighted in blue spheres. The (3S)- β -phenylalanine is highlighted in red spheres and (3S)- α -phenylalanine in highlighted in orange spheres.....	32
Figure I.15	The active site of chain A in <i>Pa</i> PAM (left) showing both (2S)- α -phenylalanine (orange) and (3S)- β -phenylalanine (red) attached to the MIO (blue) and the active site of chain B (right) showing (3S)- β -phenylalanine alone attached to the MIO. Such an arrangement of the ligands is a clear indication of a amino-group alkylation pathway. The grey mesh represents the electron density around the ligands observed at 1.2 sigma.....	33
Figure I.16	Key residues in the MIO active site of <i>Pa</i> PAM “subunit B” are highlighted in green. “Subunit A” is highlighted in magenta and “subunit C” is highlighted in cyan. The MIO is highlighted in blue. The (3S)- β -phenylalanine is highlighted in red and (2S)- α -phenylalanine in highlighted in orange. Residues from all three subunits are needed to form the active site.....	35

- Figure I.17** The active site residues of *Pa*PAM (cyan, PDB ID 3UNV) are overlaid with those of *Rt*PAL (green, PDB ID 1T6J). The left hand side shows the various hydrophobic residues that make up the binding site for the phenyl ring of the phenylalanine ligand with residue identification numbers for *Pa*PAM appearing on top of those for *Rt*PAL. Certain conserved residues are highlighted in blue boxes. The ammonia adduct attached to the methylenedioxy of the *Rt*PAL MIO (red) is positioned between the two nitrogens of the phenylalanine ligands of *Pa*PAM: (2*S*)- α -phenylalanine (orange) and (3*S*)- β -phenylalanine (red) which are attached to the MIO of *Pa*PAM (blue). Cinnamate belonging to the *Rt*PAL structure is highlighted in magenta and rests just above the active site in line with Tyr78 of *Pa*PAM.....37
- Figure I.18** Overlay of *Pa*PAM residues for the various monomers (shown in green, magenta and cyan) with bound ligands (2*S*)- α -phenylalanine (orange) and (3*S*)- β -phenylalanine (red) which are attached to the MIO (blue) (PDB ID 3UNV) and *Sg*TAM with bound inhibitor (2*S*,3*S*)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid (pink, PDB ID 2RJR) shows many charged residues around the phenyl ring in *Sg*TAM are non-polar in *Pa*PAM giving rise to the difference in substrate specificity. In addition to the ligands, the MIOs, Arg323/Arg311, Asn220/Asn205, Phe369/Phe356 and Tyr320/Tyr308 overlap one another, suggesting the control of substrate specificity is the same for both enzymes.....39
- Figure I.19** Overlay of the active site residues of *Sg*TAM with the inhibitor (2*S*,3*S*)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid (pink, PDB ID 2RJR) and the substrate mimic (3*R*)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (red, PDB ID 2RJS) show the residues in the active site are in essentially identical positions for both structures. This suggests that movement of residues is not necessary for reaction and therefore residues occupying the same positions in *Pa*PAM are aiding in the stereoslectivity of both enzymes.....41
- Figure I.20** Overlay of *Pa*PAM residues for the various monomers (shown in green, magenta and cyan) with bound ligands (2*S*)- α -phenylalanine (orange) and (3*S*)- β -phenylalanine (red) which are attached to the MIO (blue) (PDB ID 3UNV) and *Sg*TAM with bound substrate mimic (3*R*)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (red, PDB ID 2RJS) shows the depression of the para substituted phenyl ring of the substrate bound *Sg*TAM compared to the location of the un-substituted phenyl rings of *Pa*PAM. Such a movement towards the MIO suggests tight packing interactions of the inner and outer loops, which likely control the preference for aminomutase activity over lyase activity in *Pa*PAM and *Sg*TAM.....43

- Figure I.21** Overlay of “subunit A” MIO active site of PaPAM with SgTAM reveals identical trajectories of the bound ligands. The MIO, (3S)- β -phenylalanine, and (2S)- α -phenylalanine of PaPAM is highlighted in yellow. The MIO of SgTAM with bound inhibitor (2S,3S)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid (PDB ID 2RJR) is highlighted in cyan and the MIO of SgTAM with bound substrate mimic (3R)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (PDB ID 2RJS) is highlighted in green.....44
- Figure I.22** A comparison of the active site residues of *PaPAM* with (3S)- β -phenylalanine bound (green, PDB ID 3UNV) and *TcPAM* with cinnamate bound (blue, PDB ID 3NZ4) reveals conserved positions for key catalytic residues Tyr80/Tyr78, Tyr322/Tyr320 and Arg325/Arg323. Residues listed on top refer to *PaPAM* and those listed below are that of *TcPAM*.....46
- Figure I.23** Comparison of the placement of the ligands within the active sites of PaPAM (green, PDB ID 3UNV) and TcPAM (blue, PDB ID 3NZ4) shows the cinnamate in TcPAM is lifted slightly away from the MIO as compared to the (3S)- β -phenylalanine bound to the MIO of PaPAM. This suggests there is slightly more room above the ligand in TcPAM, perhaps allowing for rotation of the ligand which would allow for stereochemistry observed.....47
- Figure I.24** Overlays of *PaPAM* “subunit C” MIO active site with (3S)- β -phenylalanine bound (yellow), *TcPAM* (PDB ID 3NZ4) with cinnamate bound (blue), and *SgTAM* (PDB ID 2RJS) with (3R)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (green) bound highlight the variance in the position of the β -carbon within the propenoate moiety which is due largely to the hydrogen bonding interactions with a conserved, flexible Arg that positions the carboxylates.....48
- Figure I.25** Top: Interactions of (3S)- β -phenylalanine in *PaPAM* with Arg323 (green, PDB ID 3UNV) are compared to those of *TcPAM* with cinnamate bound to Arg325 (blue, PDB ID 3NZ4). Dark hash marks represent the hydrogen bonds to Arg325 in TcPAM as opposed to the grey hash marks representing the hydrogen bonds to Arg323 in PaPAM which allows the carboxylates of the ligands to occupy different positions. Bottom: The same as above but with space filling representations for Phe455 in PaPAM, Arg325 in TcPAM and cinnamate in TcPAM showing the collision of the cinnamate ligand to Phe455. Such a collision shows the impact Phe455 has in PaPAM in positioning the ligand away from Arg323 such that the nature of the hydrogen bonds is now different. This difference may be important in determining the stereoselectivity of the reaction.....50

Figure I.26 The active site residues of *PaPAM* with (3*S*)- β -phenylalanine (green), *TcPAM* with cinnamate (blue), and *SgTAM* with (3*R*)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (red) are overlaid with the ligand (2*S*,3*S*)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid (cyan) from *SgTAM* showing that at the location of Phe455 in *PaPAM* are non-conserved Asn residues in *TcPAM* and *SgTAM*. Adjacent to this position is a non-conserved Thr452 which pairs to Glu455 in *TcPAM* and Asn438 in *SgTAM*. These two sites may be responsible together for the stereospecificity of these aminomutases. Residues are listed in the order *TcPAM* (top), *PaPAM* (middle), and *SgTAM* (bottom). Hydrogen bonding interactions for *SgTAM* are shown as black hash lines highlighting the additional connections to the phenyl ring of the ligand. Hydrogen bonding interactions for *TcPAM* are shown in orange hash marks to show the interactions between Glu455 and Asn458 which may be important in stereoselectivity.....55

Chapter II

Figure II.1 The ligation of Coenzyme A to benzoic acid by the *Rhodospseudomonas palustris* Benzoate-Coenzyme A ligase (BadA) is an important step in producing benzyl-CoA, a substrate used biosynthetically to produce small molecules such as Taxol.....65

Figure II.2 The adenylation and thioesterification conformations are related by a large C-terminal domain movement in this class of benzoate CoA ligases. Left: CBL (N-terminal domain: red; C-terminal domain: green) with 4-chlorobenzoyl-AMP (magenta spheres) bound in the active site at the interface of the two domains. Right: Overlay of CBL with BCL_M (cyan) demonstrates the perfect overlap of the N-terminal domains compared to the large domain movement of the C-terminal domains.....67

Figure II.3 The various arrangements of the active site are highly affected by the ligand attached in CBL. Top: 4-chlorobenzoate bound showing the openness of the binding pocket towards the para position of the phenyl ring. Bottom: 4-chlorobenzoyl-AMP bound. The ATP binding pocket is just adjacent to the 4-chlorobenzoyl binding pocket. Next page: 4-chlorophenacyl CoA bound showing the CoA channel which extends to the protein surface (left side of image). Ligands are highlighted in magenta. The protein residues are shown in green. Waters appear in red. The colored mesh represents the surface area the ligand is exposed to.....68

Figure II.4 The proposed mechanism of CoA acylation in CBL includes conformational changes.....72

Figure II.5 Key residues in the benzoate binding pocket (magenta) of BCL in the adenylation conformation pack tightly against the benzoate aromatic ring (green) as evident by the surface of the binding site residues which is shown as a grey mesh.....74

Figure II.6	Crystals of <i>Rhodopseudomonas palustris</i> Benzoate-Coenzyme A Ligase.....	76
Figure II.7	Ligands used in the co-crystallization experiments of BadA.....	78
Figure II.8	Overlay of acyl-adenylate bound CBL (cyan) and acyl-adenylate bound BadA (magenta) showing on the left side the large domain shift indicative of the two separate conformations: the adenylation conformation (CBL) and the thioesterification conformation (BadA). As neither benzoic acid CoA ligase is present in both conformations they must be looked at together to gain insight into the various conformations the enzyme adopts during catalysis. The acyl-adenylate of BadA is highlighted in green spheres.....	84
Figure II.9	The switch from the adenylation conformation to the thioesterification conformation in BadA (cyan) allows Phe226 to swing out of the CoA binding channel. Residues in BCL _M are marked in green. Hydrogen bonds between the carboxylates and close residues are marked in black for BadA and orange for BCL _M . Grey mesh representing the interior surface of BCL _M shows that the CoA channel is completely blocked by Phe236.....	90
Figure II.10	Overlay of benzoic acid (green) within the active site with 2-fluorobenzoic acid (aqua), 2-toluic acid (yellow), 2-furonic acid (purple) and 2-thiophenic acid (pink).....	93
Figure II.11	The binding of benzoyl-AMP in BadA is unique among the known ligase structures. Residues (green) involved in stabilizing the benzoyl-AMP intermediate (magenta) in BadA are shown. Black hash marks denote hydrogen bonding interactions between the ligand and enzyme.....	95
Figure II.12	Overlay of residues involved in benzoyl-AMP binding in BadA (cyan) and CBL (magenta) highlight similarities in the arrangements of the residues with the exceptions of Phe226 and Lys427 in BadA. Residues listed in cyan are those of BadA and residues listed in magenta (typically appearing below those of BadA) are those of CBL.....	97
Figure II.13	Comparison of the residues (blue) involved in the acyl-adenylate of BadA (magenta) and the residues (grey) involved in binding ATP (yellow) in Human Medium-chain Acyl-coenzyme A Synthetase ACSM2A. The pocket occupied by the phosphates of ATP in Human Medium-chain Acyl-coenzyme A Synthetase ACSM2A is defined by the P-loop (grey cartoon).....	99
Figure II.14	Interactions between the hinge residues of the N- and C-terminal domain of BadA (green) with benzoyl-AMP bound in the active site (magenta) demonstrates the network of hydrogen bonds formed while BadA is in the thioesterification conformation. Black hash marks denote hydrogen bonds.....	102

Figure II.15	Overlay of benzyl CoA found in CBL (magenta) with BadA suggests amino acids relevant for CoA binding (green). Rotation of Arg250 would allow for interaction with the CoA phosphate group.....	104
---------------------	--	-----

Chapter III

Figure III.1	A) Schematic of the U1 and U6 promoters involved in RNA Polymerase II and III transcription initiation, respectively, adapted from Hernandez et. al. (7). B) Representation of the PIC of SNAPc in RNAP Polymerase III recruitment to the U6 promoter adapted from Hanzlowsky et. al. (14).....	113
Figure III.2	Schematic of the cloning for the ORF of tagless SNAP190 (1-505).....	123
Figure III.3	Schematics of the pMal-c4x, pGST_190(1-505), pSUMO/SMT3, and pMAL_SUMO_S190 (1-505) showing the relevant restriction sites.....	127
Figure III.4	Schematic of the pCDF_50/43 and pCDF_pMAL_SUMO_50/43 plasmids showing relevant restriction sites.....	130
Figure III.5	SDS Page gel. Lane 1: Crude lysate of mSNAPc. Lanes 2, 3: Crude protein elutions of mSNAPc.....	138
Figure III.6	SDS Page gel and gel filtration chromatograph of crude mSNAPc. Lane Crude: Load of mSNAPc after Ni affinity purification. Lanes 8-12: Fractions 8-12 of the chromatograph for the gel filtration shown to the left.....	139
Figure III.7	GST binding optimization of mSNAPc. Lane 1: Molecular weight standard. Lane 2: Crude mSNAPc γ 4. Lane 3: One hour binding. Lane 4: Two hours binding. Lane 5: Three hours binding. Lane 6: Four hours binding. Lanes 7-8: Overnight binding.....	141
Figure III.8	Gel of thrombin digestion optimization. Lane 0: Crude mSNAPc. Lanes 1-8: One hour increments of digestion with thrombin showing cleavage of the affinity tag.....	141
Figure III.9	Gel filtration of crude mSNAPc treated with lauryl maltoside (left) and Tween 20 (right) both showing aggregation in fraction 9 and pure protein in fraction 11.....	143
Figure III.10	Left: SDS page gel of DNase treated mSNAPc. Lane 1: Crude mSNAPc before gel filtration. Lanes 2-10: Fractions 9 – 17 of the resulting gel filtration. Right: Superdex 200 gel filtration of crude mSNAPc before (top) and after (bottom) treatment with DNase. Fraction 9 represents the void peak. Fraction 11 represents pure mSNAPc protein.....	144

- Figure III.11** Gel filtration of 4 combined previously run gel filtrations of mSNAPc representing 25 total purifications of 150 total liters of cell culture.....145
- Figure III.12** Left: Needle of mSNAPc grown in 0.1M HEPES, pH 8.0, 0.05M MgCl₂, 0.2M NaCl, and 8% PEG mme 5000. Right: The same needles as on the left under polarized light.....146
- Figure III.13** SDS Page gel of crude mSNAPc. Lane Mw: Molecular weight standard. Lanes 1 and 2: Original Purification of mSNAPc. Lanes 3-5: Different redundant purifications of mSNAPc with 1M total KCl in HEG buffers. Lane 6: Purification of mSNAPc with 1M KCl total in HEG buffers plus lauryl maltoside in the wash buffer. SNAP19 is running with the salt front of the gel.....147
- Figure III.14** Gel filtration of 6 combined purifications of crude mSNAPc purified as described above. Fraction 13 represents the void. Fraction 18 represents pure protein.....148
- Figure III.15** SDS Page gel of co-expressed His-tagged SNAP50 and SNAP43. Lane MW: Molecular Weight. Lane Crude: Crude lysate. Lanes Wash 1 and Wash2: 20 mM imidazole containing buffer wash of Ni resin. Lanes E1-E4: 250 mM imidazole containing buffer elutions of Ni resin. Lane γ 4: Sample of pure mSNAPc showing the location of the SNAP50 and SNAP43 bands.....149
- Figure III.16** SDS Page gel of co-expressed His-tagged SNAP50, SNAP43, and N-terminal GST tagged SNAP190 (1-505). Lane MW: Molecular Weight. Lane Crude: Crude lysate. Lane Wash 1: 20 mM imidazole containing buffer wash of Ni resin. Lanes E1-E4: 250 mM imidazole containing buffer elutions of Ni resin. Lane γ 4: Sample of pure mSNAPc showing the location of the SNAP50 and SNAP43 bands.....150
- Figure III.17** Top: SDS Page gel of co-expressed SNAP50 (N-terminal His tag), SNAP43, SNAP190 (1-505 with C-terminal GST tag) and SNAP19. Lanes 1 and 10: Molecular weight standard. Lane 2: FT of wash buffer. Lane 3: FT of wash buffer plus 1% lauryl maltoside. Lanes 4-8: Ni NTA Elutions of crude protein complex. Lane 8: Ni NTA resin after elution. Bottom: SDS Page gel of co-expressed SNAP50 (N-terminal His tag), SNAP43, SNAP190 (1-505 with C-terminal GST tag) and SNAP19 further purified after Ni purification. Lane 1: Molecular weight standard. Lane 2: Crude complex that flowed through the GST resin. Lane 3: Complex that was cleaved from resin after thrombin digestion.....151
- Figure III.18** Comparison of the pMAL_SUMO_S50/S43 (PS_S50) plasmid's promoter region (top) with that of the untagged original pCDF_S50 promoter region (middle). Differences are highlighted in green. The corrected pMAL_SUMO_S50/S43 plasmid (bottom) as determined by sequencing.....154

Figure III.19 The open reading frame of SNAP19. Circles highlight the leucine zipper motif. The glutamic acid region is underlined.....155

Figure III.20 Cartoon representations of the different delta constructs. Pink: Myb domain. Yellow: SNAP190-SNAP50 interacting region. Red: Thrombin cleavable linker.....156

Chapter IV

Figure IV.1 Top: Crystal structure of a human TBP core domain (blue)-human TFIIB core domain (green) complex bound to an extended, modified adenoviral major late promoter (orange) (8) (PDB ID 1C9B). Bottom: a yeast Brf1 (blue)-TBP (green)-DNA (orange) ternary complex (9) (PDB ID 1NGM). Aligning the TBP core domains gives the relative orientation of the other segments when designing fusion constructs.....162

Figure IV.2 Cartoon representation of the ORFs for several different TF fusion constructs. The black bars represent the locations of poly-histidine affinity tags.....163

Figure IV.3 SDS Page gel of a typical TF8 purification. Lane 1: Lysate pellet. Lane 2: Lysate supernatant. Lane 3: Ni NTA resin with protein bound. Lane 4: Elution of TF1. Lane 5: Molecular weight standard.....169

Figure IV.4 Chromatograph of HiTrapHeparin HP of TF8. Fractions 1-17 represent the load. Fractions 21-36 represent the salt gradient from 400-800 mM NaCl. Fraction 33 contains the major fraction of pure TF8.....171

Figure IV.5 Top: Crystals of TF1 annealed to SymSelex2. Bottom: Crystals of TF8 annealed to SymSelex2.....174

LIST OF ABBREVIATIONS

Amino Acids

Ala, A	Alanine
Arg, R	Arginine
Asn, N	Asparagine
Asp, D	Aspartic acid
Cys, C	Cysteine
Gln, Q	Glutamine
Glu, E	Glutamic acid
Gly, G	Glycine
His, H	Histidine
Ile, I	Isoleucine
Leu, L	Leucine
Lys, K	Lysine
Met, M	Methionine
Phe, F	Phenylalanine
Pro, P	Proline
Ser, S	Serine
Thr, T	Threonine
Trp, W	Tryptophan
Tyr, Y	Tyrosine
Val, V	Valine

Other Symbols and Abbreviations

Å	Ångström
°C	degrees Celsius
μM	micromolar
μL	microliter
aa	amino acid
ACSM2A	Human Medium-chain Acyl-coenzyme A Synthetase ACSM2A
AMP	adenosine monophosphate
APS	advanced photon source
ATP	adenosine triphosphate
<i>Av</i> PAL	<i>Anabaena Variabilis</i> ATCC 29413 phenylalanine ammonia lyase
BadA	benzoate coenzyme a ligase from <i>Rhodospseudomonas palustris</i>
BCL _M	<i>Burkholderia xenovorans</i> LB400 benzoate-coenzyme A ligase
BDP1, B''	B double prime
BLAST	Basic Local Alignment Search Tool
bp	base pair
BRF2	TFIIB Related Factor
B.rhiz	<i>Burkholderia rhizoxinica</i>
BSA	bovine serum albumin
B.subtl	<i>Bacillus Subtlis</i>
CBL	4-Chlorobenzoate:Coenzyme A Ligase
CCP4	Collaborative computational project, number 4
CoA	Coenzyme A

C-terminal	carboxy terminal
DBTNBT	N-debenzoyl-2'-deoxytaxol N-benzoyltransferase
DEAE	DiEthylAminoEthane
DNA	deoxyribonucleic acid
DSE	distal sequence element
DTT	dithiothreitol
<i>E. Coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
g	gram
GST	glutathione S-transferase
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
K	kilo, one thousand
KCl	potassium chloride
KDa	kilodalton
K.pneu	<i>Klebsiella pneumoniae</i> 342
L	liter
LB	luria broth
M	molar
MBP	maltose binding protein
mg	milligram
MgCl ₂	magnesium chloride
MIO	4-methylideneimidazole-5-one

mL	milliliter
mM	millimolar
mm	millimeter
MME	monomethyl ether
mRNA	messenger RNA
mSNAPc	small nuclear RNA activating protein
<i>NpPAL</i>	<i>Nostoc punctiforme</i> ATCC 29133 phenylalanine ammonia lyase
N-terminal	amino terminal
MW	molecular weight
MWCO	molecular weight cut off
NaCl	sodium chloride
NCBI	National Center for Biotechnology Information
Ni	nickel
NTA	nitrilotriacetic acid
OIR	Oct1 interacting region
OD	ocular density
ON	over night
ORF	open reading frame
<i>PaPAM</i>	<i>Pantoea agglomerans</i> phenylalanine aminomutase
PAM	phenylalanine aminomutase
<i>PcPAL</i>	<i>Petroselinum crispum</i> phenylalanine ammonia lyase
PCR	polymerase chain reaction
PDB	protein data bank

PEG	polyethylene glycol
PIC	pre-initiation complex
PMSF	phenylmethanesulfonylfluoride
Pol I, Pol II, Pol III	RNA polymerase I, II or III
<i>Pp</i> HAL	<i>Pseudomonas putida</i> histidine ammonia lyase
PSE	proximal sequence element
RCSB	Research Collaboratory for Structural Bioinformatics
R-factor	Reliability factor
RMSD	root mean square deviation
RNA	ribonucleic acid
RNAP II or III	RNA polymerase II or III
RPM	rotations per minute
rRNA	ribosomal RNA
<i>Rs</i> TAL	<i>Rhodobacter sphaeroides</i> tyrosine ammonia lyase
<i>Rt</i> PAL	<i>Rhodospiridium toruloides</i> phenylalanine ammonia lyase
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SER	surface entropy reduction
<i>Sg</i> TAM	<i>SgcC4</i> L-tyrosine 2,3-aminomutase
SMT3	<i>Saccharomyces cerevisiae</i> ubiquitin-like protein
snRNA	small nuclear RNA
<i>S.marit</i>	<i>Streptomyces maritimus</i> EncP
SUMO	small ubiquitin-like modifier
TBP	TATA Binding Protein

<i>TcPAM</i>	<i>Taxus canadensis</i> phenylalanine aminomutase
TF8	TFIIIB Brf1-TBP triple fusion
TF	transcription factor
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol
tRNA	transfer ribonucleic acid
V. bact	<i>Vibrionales bacterium</i> SWAT-3

Chapter 1: PaPAM

I.1. Background

I.1.1. Co-opting biosynthetic pathways

Living organisms are able to produce a variety of bioactive organic compounds that represent a wealth of potential drugs. However, it is often impractical to harvest these compounds directly due to the small amount of each compound produced and/or the rarity of the source producing it. Traditional organic synthetic methods for producing these compounds can be expensive; resulting in low yields and the production of large amounts of hazardous waste, while obtaining the correct stereochemistry can be elusive. Biosynthetic methods provide the opportunity of addressing these issues. In a biosynthetic strategy, recombinant enzymes either from the biosynthetic pathway of the target organism or related species are used *in vitro* or *in vivo* to transform precursor molecules with high fidelity and enantiomeric purity.

To aid in designing new biosynthetic pathways the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (<http://www.enzyme-database.org>) has to date over 5300 different enzymes in its inventory divided into classes based on their overall transformations. Programs such as the PathPred Program (<http://www.genome.jp/tools/pathpred/>) are also being developed to take target substrates and design novel biosynthetic pathways using such databases (1).

Biosynthetic methods additionally open up the possibility of replacing enzymes within a given pathway with ones that are capable of performing alternate chemical transformations, thus producing a wider range of molecules. Inherent to this strategy is the use of alternate enzymes from a variety of sources such as plants and bacteria (2). Ideally, a single enzyme capable of

producing multiple derivatives would be used. These enzymes might occur naturally, but most enzymes function in a way that is highly substrate specific.

X-ray crystallography can address this and other issues. If the structure of the enzyme is known, specific amino acids within its sequence can be logically altered allowing for broader substrate specificity (3). Crystal structures can also give clues to the type of mechanism used by the enzyme to carry out the reaction by providing a snap-shot view of the active site. At the same time the shape and character of the active site can suggest why an enzyme turns over one product and not another. How the enzyme enforces stereoselectivity might also be understood and in some cases switched to other stereochemical configurations. It is therefore desirable to obtain X-ray crystallographic structures of these enzymes.

I.1.2. Phenylalanine Aminomutases

Phenylalanine aminomutases (PAMs) remove the amino group from naturally occurring L-phenylalanine to produce β -phenylalanine derivatives (**Figure I.1**) (4-6). This transformation has important industrial implications in the production of chiral phenylalanines (2, 7) as well as in the production of the antibiotic Andrimid (4), and the anticancer drug Taxol (8).

The antibiotic Andrimid is produced by the pathogenic Gram-negative bacteria *Pantoea agglomerans* (*Pa*) (49, 50). Andrimid inhibits bacterial acetyl-CoA carboxylase (ACC) which is involved in fatty acid biosynthesis, preventing cell growth. As with any antibiotic there is a risk of bacterial resistance which makes the production of derivatives appealing; a challenge that X-ray crystallography can address.

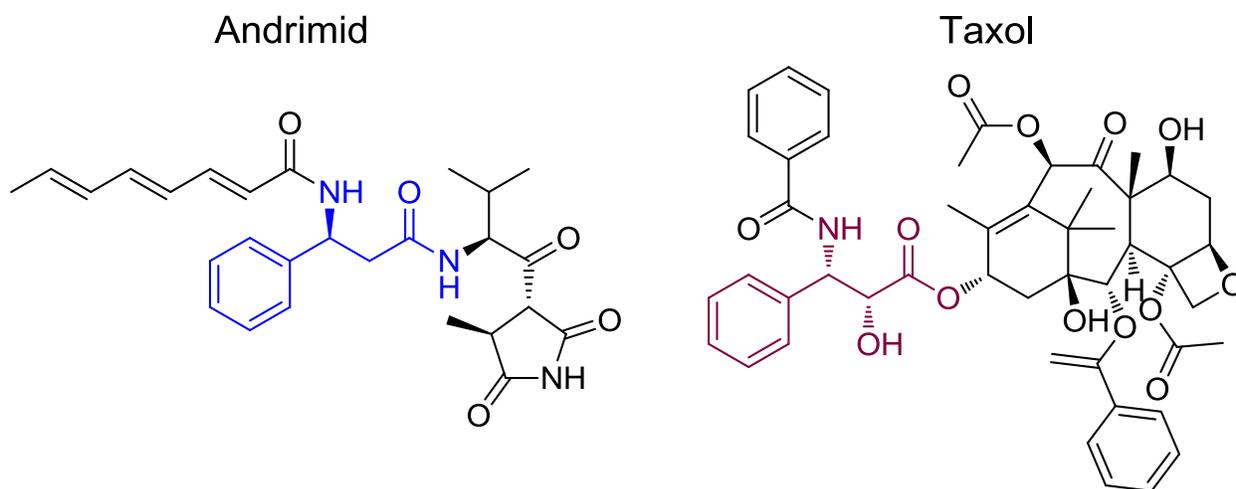
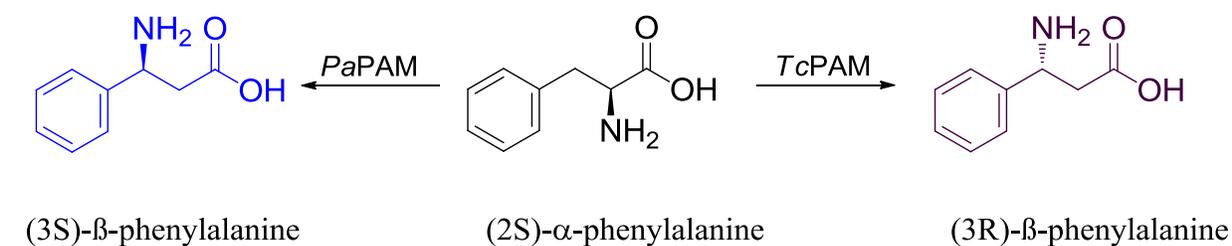


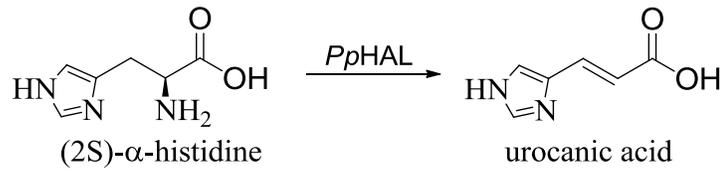
Figure I.1 Top: The conversion of (2S)-α-phenylalanine to (3R)-β-phenylalanine (purple) by *Taxus canadensis* phenylalanine aminomutase (*TcPAM*) and to (3S)-β-phenylalanine (blue) by *Pantoea agglomerans* phenylalanine aminomutase (*PaPAM*). Bottom: The antibiotic Andrimid and anti-cancer drug Taxol both contain a β-phenylalanine component (highlighted in blue and purple) and contain phenylalanine aminomutases within their biosynthetic pathways (*TcPAM* and *PaPAM*, respectively). For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

Paclitaxel (Taxol) is a diterpene alkaloid that was first isolated from the bark of the Pacific yew (*Taxus brevifolia*) in 1967 by Wall and Wani of the U.S. National Cancer Institute. It is one of the most popular anti-cancer drugs known to date. It has been found to be effective in treating a variety of cancers including ovarian and breast cancers. It is classified as an anti-mitotic drug since its mode of action is to stabilize microtubules found within the mitotic spindle, inhibiting eukaryotic cell division which leads to mitotic arrest and cell death (9). It is thought to do this by directly binding to the β subunit of tubulin (10, Brookhaven Protein Data Bank ID 1tub, see also PDB ID 1JFF by Lowe J, Li H, Downing KH and Nogales E). Taxol consists of a core taxane ring, a β -amino acid containing side chain, and has 12 stereogenic centers making it a challenging synthetic target. Derivatives that might address toxicity and drug resistance in patients are difficult to produce (11). Practical concerns stemming from the mass farming of Pacific yew trees for their bark led to the total organic synthesis of Taxol by several research groups, the first being those of Holton (12, 13) and Nicolaou (14) in 1994. Next came semi-synthetic methods using 10-deacetylbaccatin III (10-DAB) which can be extracted from the leaves and twigs of the European yew tree (14) However, biosynthesis has proven to be the most efficient route of production (16-18). Today, Bristol-Myers Squibb utilizes plant cell fermentation technology (PCF) to produce Taxol in large quantities from plant cell cultures of *Taxus spp*, eliminating the need to produce tens of thousands of pounds of hazardous materials annually (19). The use of PAMs to produce Taxol derivatives is already being investigated (20). Knowledge of the crystal structure will aid in producing potentially substrate promiscuous mutants.

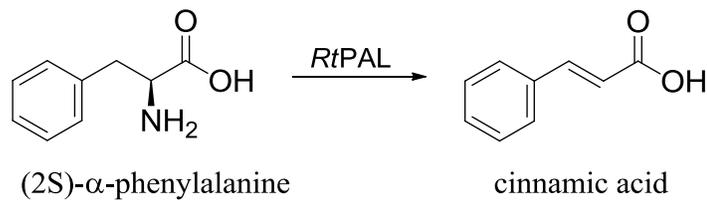
I.1.3. Ammonia Lyase Mechanistic Studies

PAMs represent a subclass of aminomutases in the class I lyase-like family. A lyase is defined generally as any enzyme that breaks chemical bonds. PAMs are closely related to tyrosine aminomutases (TAMs), tyrosine ammonia lyases (TALs), phenylalanine ammonia lyases (PALs) (21), and histidine ammonia lyases (HALs) (22, 23) (**Figure I.2**). The ammonia lyases were the first to be characterized and convert their amino acid precursors to unsaturated acids by removal of an α -amino group. The aminomutases differ in that they shuffle the α -amino group to the β position. As they are closely related in structure, active site and sequence, it is likely the same mechanism guides the reactions of this class of enzymes (22, 23).

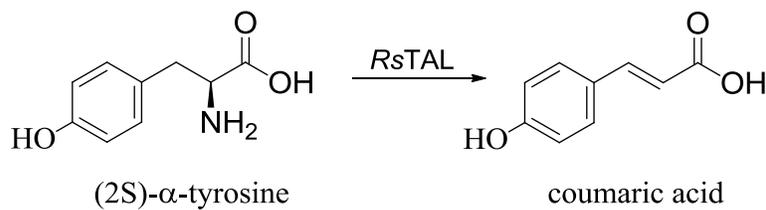
Histadine ammonia Lyases (HAL)



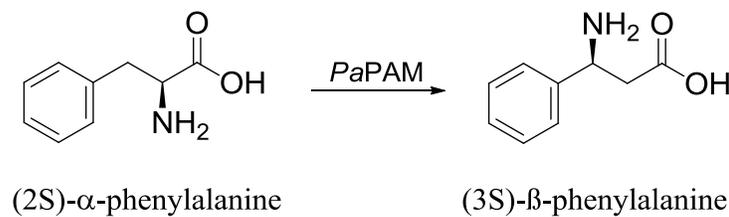
Phenylalanine ammonia Lyases (PAL)



Tyrosine ammonia Lyases (TAL)



Phenylalanine aminomutases (PAM)



Tyrosine aminomutases (TAM)

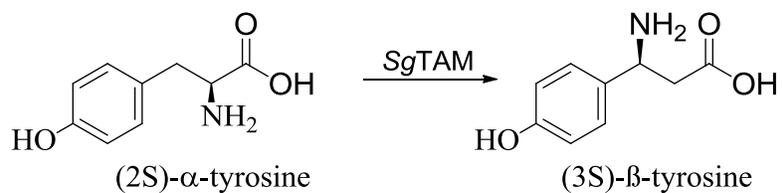


Figure I.2 PAMs are a member of the class I lyase-like family perform similar chemical transformations and include tyrosine aminomutases (TAMs), tyrosine ammonia lyases (TALs), phenylalanine ammonia lyases (PALs), and histidine ammonia lyases (HALs). Specific examples of enzymes that perform these transformations are shown above the arrows.

Before any structural data was available, it was theorized that the ammonia lyases removed the amino group through an amino-group alkylation pathway via a dehydroalanine moiety (24). Upon their subsequent discovery, it was proposed that the aminomutases followed a similar mechanism with the amino group rebounding to the unsaturated acid to produce the β -amino acid derivatives.

The first of this family to be structurally identified was *Pseudomonas putida* histidine ammonia lyase (*Pp*HAL) (22). The structure of *Pp*HAL proved that members of the class I lyase-like family contain a 4-methylideneimidazole-5-one (MIO) prosthetic group which catalyses their respective reactions (25) and not dehydroalanine. The MIO is formed auto-catalytically from a chain of three amino acids post-translation, similar to the chromophore first found in the *Aequorea victoria* Green Fluorescent Protein (26). Common to the class I lyase-like family, the sequence alanine-serine-glycine forms the MIO. For *Pa*PAM the MIO is formed from the sequence threonine-serine-glycine (**Figure I.3**).

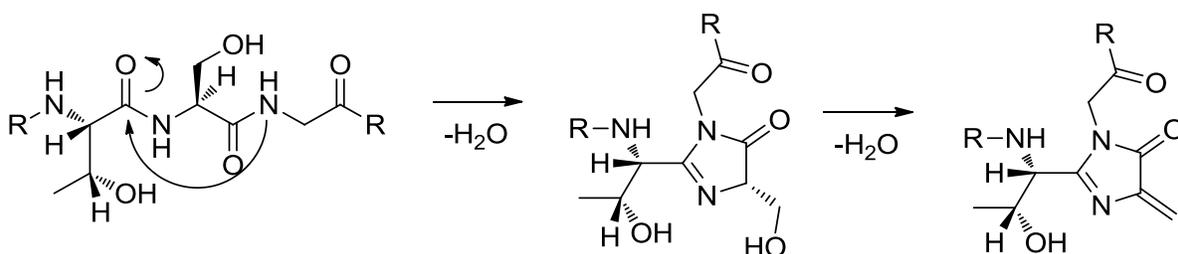
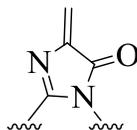


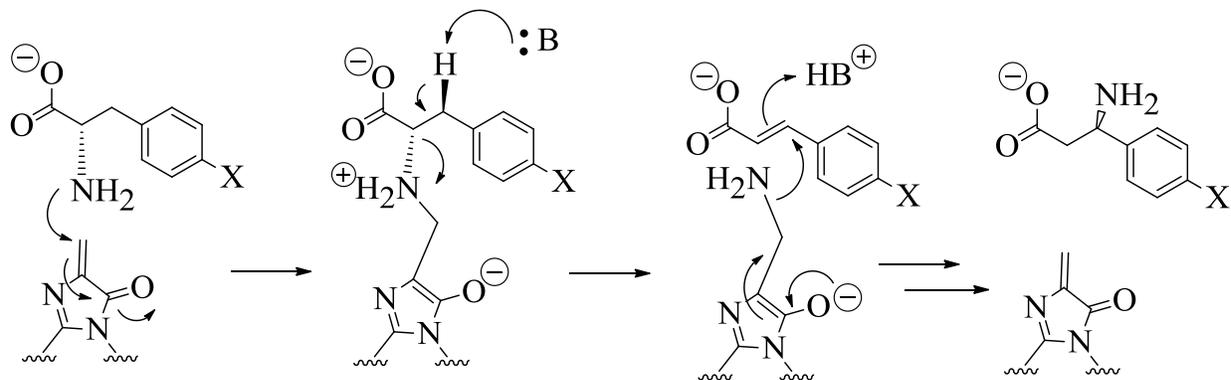
Figure I.3 The autocatalytic formation of the MIO in *Pa*PAM from the sequence threonine-serine-glycine is adapted from that of HAL (27).

There is much controversy concerning the mechanism of these reactions as they relate to the MIO. In one scheme, the MIO participates in an amino-group alkylation pathway by first removing the amino group from the amino acid substrate via nucleophilic attack of the methyldene group of the MIO by the amino group and subsequently replacing it on a *trans*-aryl acrylate intermediate stereoselectively (**Figure I.4**) (28). This mechanism is an adaptation of the amino-group alkylation pathway proposed when the MIO was believed to be dehydroalanine. It is supported by observed kinetic isotope effects, the ordered release of products, and intermediates and products characterized from these enzymes (29 and references within). Despite its simplicity, there has been some doubt as to the validity of this mechanism because the pKa for the benzylic proton that must be removed is thought to be high (>40) (30). In the alternative scheme, the MIO instead participates in a Friedel-Craft-type reaction that activates the benzylic proton by nucleophilic attack of the MIO by the ortho position of the aromatic group, thus catalyzing the formation of the *trans*-aryl acrylate intermediate. This intermediate then reacts with free ammonia to form the final product (31). Evidence to support this mechanism, specifically in HALs, include isotope effects concerning the hydrogens present on the aromatic ring. Such an effect suggests that the hydrogens are directly involved in the reaction (29 and references within). Additionally designed small molecule systems meant to mimic the MIO's activity in PALs and the activity of substituted alternate substrates in HALs lend support to a Friedel-Craft-type mechanism (29 and references within).

4-methylideneimidazole-5-one: MIO



Amino-group alkylation mechanism:



Friedel-Crafts type mechanism:

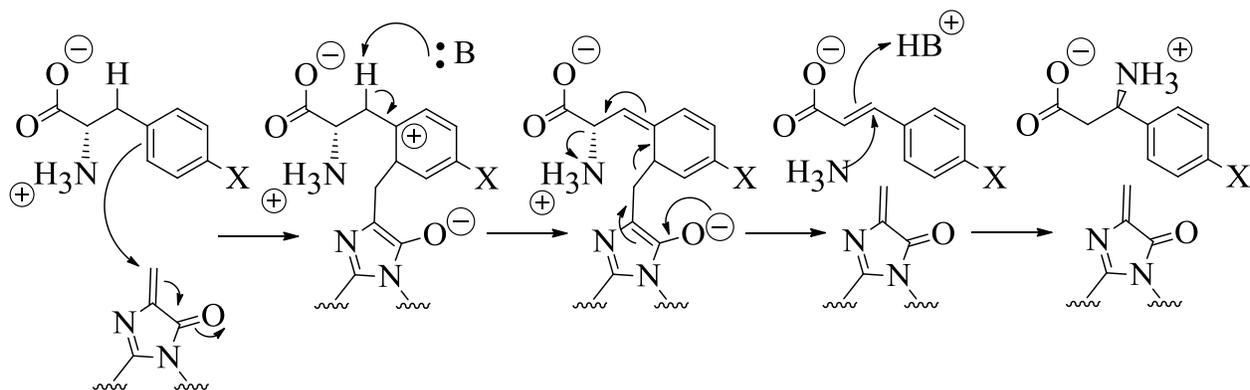


Figure I.4 Top: Structure of the MIO. Bottom: Two proposed mechanisms for the conversion of substrate to product in a generic aminomutase. X represents hydrogen in the case of phenylalanine and a hydroxyl group in the case of tyrosine. For a generic ammonia ligase the reaction would stop with the production of *trans*-cinnamate in the case of PALs or *trans*-coumerate in the case of TALs.

Understanding the mechanism of action is critical if these enzymes are to be fully exploited for biocatalysis. Since the MIO is ubiquitous to both mechanisms, it falls to the residues in the active site as well as the tertiary structure of the protein to give clues as to which mechanism, if either, is correct. PAMs, TAMs, TALs, HALs and PALs are homologs, having common domain folds and high levels of amino acid sequence similarity (25). The key differences in the overall reactions are the preference for phenylalanine or histidine versus tyrosine as the initial substrate as well as the final product being *trans*-cinnamic acid (when phenylalanine is the substrate), *trans*-coumarate (when tyrosine is the substrate), *trans*-urocanate (when histidine is the substrate) or a β -phenylalanine or β -tyrosine derivative (in the case of the aminomutases).

I.1.4. Structures of MIO-containing enzymes

As previously stated, *Pp*HAL “histidase” was the first of the class I lyase-like family members whose structure was determined revealing the presence of the MIO group. It catalyzes the removal of ammonia from histidine to form *trans*-urocanate, a compound which is believed to offer sun protection in human skin. In humans deficiencies of histidase causes histidinemia (22 and references within). To aid in comparing the structures of members of the class I lyase-like family, an alignment of some of the sequences for those members who are mentioned below is given in **Table I.1**. Some of the highly conserved residues which are of mechanistic significance are highlighted

Table I.1 Sequence alignment of class I lyase-like family members using CLUSTAL 2.1 (32).

Some key residues have been highlighted. Proteins are identified by their PDB ID codes followed by the abbreviations as follows: *Pp*HAL: *Pseudomonas putida* histidine ammonia lyase; *Rs*TAL: *Rhodobacter sphaeroides* tyrosine ammonia lyase; *Pa*PAM: *Pantoea agglomerans* phenylalanine aminomutase; *Sg*TAM: *SgcC4* L-tyrosine 2,3-aminomutase; *Pc*PAL: *Petroselinum crispum* phenylalanine ammonia lyase; *Tc*PAM: *Taxus canadensis* phenylalanine aminomutase; *Rt*PAL: *Rhodospiridium toruloides* phenylalanine ammonia lyase; *Av*PAL: *Anabaena Variabilis* ATCC 29413 phenylalanine ammonia lyase; *Np*PAL: *Nostoc punctiforme* ATCC 29133 phenylalanine ammonia lyase

1GKM_PpHAL	-----TELTLKPG	8
2O6Y_RsTAL	-----MLAMSPP	7
3UNV_PaPAM	-----MSIVNESGSQPVVSRDETLTSLQIERTS	26
2OHY_SgTAM	-----GS-----MALTQVETEI	12
1W27_PcPAL	---MENGNGATTNGHVNGNGMDFCMKTEDPLYWGGIAAEAMTG---SHLDEVKMKVAEYRK	54
3NZ4_TcPAM	----MGFAVESRSHVKD-----ILGLINT-----FNEVKKITVD---	30
1T6P_RtPAL	---MAPSLDSISHSFANGVASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAAPT	57
2NYN_AvPAL	MKTLTSLQVSKTSS-----QQFSFTGNSS	23
2NYF_NpPAL	MNITSLQQNITRS-----WQIPFTNSSD	23
1GKM_PpHAL	-----TLTLAQLRAIHAA---PVRLQLDASAAPAIASVACVEQIIAEDRTAYGINT	57
2O6Y_RsTAL	KPAVELDRHIDLDAQAHAVASG---GARIVLAPPARDRCRASEARLGAVIREARHVYGLTT	64
3UNV_PaPAM	FHISS-GKDISLEEIARAARD---HQPVTTLHDEVVNRVTRSRSILESMVSDERVIYGVNT	82
2OHY_SgTAM	VPVSVGDGETLTVEAVRRVAEE---RATVDVPAESIAKAQKSREIFEGIAEQNIPIYGVTT	69
1W27_PcPAL	PVVKLGGETLTISQVAAISARDGSGVTVELSEAARAGVKASSDWVMSMNGKGTDSYGVTT	114
3NZ4_TcPAM	-----GTTPIITVAHVAALARRHDVKVALEA-EQCRARVETCSSSWVQRKAEDGADIYGVTT	84
1T6P_RtPAL	STLELDGYSNLGADVSAARK-GRPVRVKDSDEIRSKIDKSVEFLRSQLS--MSVYGVTT	114
2NYN_AvPAL	ANVIIGNQKLTINDVARVARN-GTLVSLTNNTDILQGIQASCDYINNAVESGEPIYGVTS	82
2NYF_NpPAL	SIVTVGDRNLTIDEVVNVARH-GTQVRLTDNADVIRGVQASCDYINNAVETAQPIYGVTS	82
	: : : . . **::	
1GKM_PpHAL	GFGLLASTRIASHDLENLQRSLVLSHAAGIG-----APLDDDLVRLIMVLK	103
2O6Y_RsTAL	GFGPLANRLISGENVRTLQANLVHHLASGVG-----PVLDTWTARAMVLAR	110
3UNV_PaPAM	SMGGFVNYIVPIAKASELQNNLINAVATNVG-----KYFDDTTVRATMLAR	128
2OHY_SgTAM	GYGEMIYMQVDKSKEVELQTNLVRSHSAGVG-----PLFAEDEARAIVAAR	115
1W27_PcPAL	GFGATSHRRTK--QGALQKELIRFLNAGIFGNGSD-----NTLPHSATRAAMLVR	163
3NZ4_TcPAM	GFGACSSRRTN--QLSELQESLIRCLLAGVFTKGCASS-----VDELTPATVTRSAMLLR	136
1T6P_RtPAL	GFGGSADTRTE--DAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIR	172
2NYN_AvPAL	GFGGMANVAISREQASELQTNLVWFLKTGAG-----NKLPLADVRAAMLLR	128
2NYF_NpPAL	GFGGMADVVISREQAAELQTNLIWFLKSGAG-----NKLPLADVRAAMLLR	128
	. * . ** * : . : . * : :	
1GKM_PpHAL	INLSRFGSGIRRKVIDALIALVNAEVYPHIPLKGSVGASGDLAPLAHMSLVLLGEGKAR	163
2O6Y_RsTAL	LVSIAQGASGASEGTIARLIDLNLSELAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFL	170
3UNV_PaPAM	IVLSLRGNSAISIVNFKKLIEIYNQGIVPCIPEKGSGLTSGDLGPLAAIALVCTGQWKAR	188
2OHY_SgTAM	LNTLAKGHS AVRPIILERLAQYLNEGITPAIPEIGSLGASGDLAPLSHVASTLIGEGYVL	175
1W27_PcPAL	INTLLQGYS GIRFEILEAITKFLNQNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNS-	222
3NZ4_TcPAM	LNSFTYGC SGIRWEVMEALEKLLNSNVSPKVPLRGSVSASGDLIPLAYIAGLLIGKPSV-	195
1T6P_RtPAL	VNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSK	232
2NYN_AvPAL	ANSHMRGASGIRLELIKRMIEIFLNAGVTPYVYEFGSIGASGDLVPLSYITGSLIGLDP-	187
2NYF_NpPAL	ANSHLYGASGIRLELIQRIETFLNAGVTPHYVYEFGSIGASGDLVPLSYITGALIGLDP-	187
	: * * : : * : * : ** : ***** * : *	

Table I.1 (cont'd)

1GKM_PpHAL	-----YKGQWLSATEALAVAGLEP--LTLAAKEGLLALLNGTQASTAYALRGLFYAEDLY	215
2O6Y_RsTAL	D-----RDGTRLDGAEGLRRGRQLP--LDLSHRDALALVNGTSAMTGIALVNAHACRHLG	223
3UNV_PaPAM	-----YQGEQMSGAMALEKAGISP--MELSFKEGLALINGTSAMVGLGVLLYDEVKRLF	240
2OHY_SgTAM	-----RDGRPVETAQVLAERGI EP--LELRFKEGLALINGTSGMTGLGSLVVGRALEQA	227
1W27_PcPAL	--KAVGPTGVILSPEEAFKLAGVEGGFFELQPK EGLALVNGTAVGSGMASMVLFEANILA	280
3NZ4_TcPAM	--IARIGDDVEVPAP EALS RVGLRP--FKLQAKEGLALVNGTSFATALASTVMYDANVLL	251
1T6P_RtPAL	VHVHHEGKEKILYAREAMALFNLEP--VVLGPKEGLGLVNGTAVSASMATLALHDAHMLS	290
2NYN_AvPAL	--FKVDFNGKEMDAPTALRQLNLSP--LTL LPK EGLAMMNGTSVMTGIAANC VYDTQILT	243
2NYF_NpPAL	--FTVDFD GKEMDAVTALSRLGLPK--LQLQPK EGLAMMNGTSVMTGIAANC VYDAKVL L	243
	: : : . * : : * : : * * . .	
1GKM_PpHAL	AAAIACGGLSVEAVLGSRS PFDARIHEAR-GQRGQIDTAACFRDLLGDSSEVS---LSHK	271
2O6Y_RsTAL	NWAVALTALLAECLRGRTEAWAAALS DLR-PHPGQKDAARLRARVDGSARVVRHVIAER	282
3UNV_PaPAM	DTYLTVTSL SIEGLHGKTKPFEP AVHRMK-PHQGQLEVATTIWETLADSSLAVNEHEVEK	299
2OHY_SgTAM	QQAEIVTALLIEAVRGSTSPFLAEGHDIARPHEGQIDTAANMRALMRGSGLTVEHADLRR	287
1W27_PcPAL	VLAEVMSAIFAEVMQ GKPEFTDHLTHK LK-HHPGQIEAAAIMEHILDGSAYV KAAQKLHE	339
3NZ4_TcPAM	LLVETLCGMFCEVIFGREEFAHPLIHKVK-PHPGQIESAE LLEWLLRSSPFQDLSREYYS	310
1T6P_RtPAL	LLSQSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLL EGSRFVHHEEEVK	350
2NYN_AvPAL	AIAMGVHALDIQALNGTNQSFHPFIHNSK-PHPGQLWAADQMI SLLANSQ LVRDEL DGKH	302
2NYF_NpPAL	ALTMGVHALAIQGLYGTNQS FHPFIHQCK-PHPGQLWTADQMF SLLKDS SSVREELDGKH	302
	. : : : * : * * : : . *	
1GKM_PpHAL	NADK-----VQDPYSLRCQPQVMGACLTQLRQAAEVLGIEAN-AVSDNPLV	316
2O6Y_RsTAL	RLDAGDIG-----TEPEAGQDAYS LRCAPQVLGAGFDTLAWHDRVLTIELN-AVTDNP-V	335
3UNV_PaPAM	LIAEEMDG--LVKASNHQIEDAYSIRCTPQILGPVADTLKNIKQTLTNELN-SSNDNP--	354
2OHY_SgTAM	ELQKDKEAGKDVQRSEIYLQKAYS LRAIPQVVGAVRDTLYHARHKLRIELN-SANDNP--	344
1W27_PcPAL	MDPLQ-----KPKQDRYALRTSPQWLGPQIEVIRSSTKMIEREIN-SVNDNP--	385
3NZ4_TcPAM	IDK LK-----KPKQDRYALRSSPQWLAPLVQTIRDATTTVETEVEVN-SANDNP--	356
1T6P_RtPAL	VKDDEG-----ILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNP--	398
2NYN_AvPAL	DYRDH-----ELIQDRYSLRCLPQYLGPIVDGISQIAKQIEIEIN-SVTDNP--	348
2NYF_NpPAL	EYRGK-----DLIQDRYSLRCLAQFIGPIVDGVSEITKQIEVEMN-SVTDNP--	348
	: . * : * . * : . . : : * . : . * * *	
1GKM_PpHAL	FAAEG--DVISGGNFHAEPVAMAADNLALAI AEIGSLSERRISLMMDKHMSQ-LPPFLVE	373
2O6Y_RsTAL	FPPDGSVPALHGGNFMGQHVALTSDALATAVTVLAGLAERQIARLTDERLNRGLPPFLHR	395
3UNV_PaPAM	LIDQTTEE VFNHGHFHGQYVSMAMDHLNIALVTMMNLANRRIDRFMDKSNSNGLPPFLCA	414
2OHY_SgTAM	LFFEG-KEIFHG ANFHGQPIAFAMDFVTIALTQLGVLAERQINRVLNRHLSYGLPEFLVS	403
1W27_PcPAL	LIDVSRNKAIHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNL SG	445
3NZ4_TcPAM	IIDHANDRALHGANFQGS AVGFYMDYVRIAVAGLGKLLFAQFTELMIEYYSNGLPGNL SL	416
1T6P_RtPAL	LIDVENKTSHHGGNFQAAA VANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLP SCLAA	458
2NYN_AvPAL	LIDVDNQASYHGGNFLGQYVGMGMDHLRY YIGLLAKHLDVQIALLASPEFSNGLPPSLLG	408
2NYF_NpPAL	LIDVENQVSYHGGNFLGQYVGMTDRLRY YIGLLAKHIDVQIALLVSPEFSNGLPPSLVG	408
	: . : * . : . : : : : . . * * *	

Table I.1 (cont'd)

1GKM_PpHAL	N-GGVNSG-FMIAQVTAALASENKALSHPHSVDSLP-TSANQEDHVSMAPAAGKRLWEM	430
2O6Y_RsTAL	GPAGLNSG-FMGAQVTATALLAEMRATG-PASIHSSIS-TNAANQDVVSLGTIAARLCREK	452
3UNV_PaPAM	ENAGLRLG-LMGGQFMTASITAESRASCMPMSIQSLS-TTGDFQDIVSFGLVAARRVREQ	472
2OHY_SgTAM	GDPGLHSG-FAGAQPATALVAENRT-IGPASTQSVS-SNGDNQDVVSMGLISARNARRV	460
1W27_PcPAL	GRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQS-AEQHNQDVNSLGLISSRKTSEA	504
3NZ4_TcPAM	GPDLSDYGLKGLDIAMAAYSSELQYLANPVTTHVHS-AEQHNQDINSLALISARKTDEA	475
1T6P_RtPAL	E-DPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQP-AEMANQAVNSLALISARRTTES	516
2NYN_AvPAL	NRERKVNMGKGLQICGNSIMPLLFYGNISADRFPHTAEQFNQNINSQGYTSATLARRS	468
2NYF_NpPAL	NSDRKVNMGKGLQISGNSIMPLLSFYGNLADRFPHTAEQFNQNINSQGYISANLTRRS	468
	: : . . : . : * . :.	.
1GKM_PpHAL	AENTRGVLAIEWLGACQGLDLRKG--LK-TSAKLEKARQALR-----	469
2O6Y_RsTAL	IDRWAEILAILALCLAQAAELRCGSLDGVSPAGKKLVQALR-----	494
3UNV_PaPAM	LKNLKYVFSFELLCACQAVDIRG---TAGLSKRTRALYDKTR-----	511
2OHY_SgTAM	LSNNNKILAVEYLAAAQAVDISGR--FDGLSPAATYEAVER-----	500
1W27_PcPAL	VEILKLMSTTFLVGLCQAIDLRHLEENLKSTVKNTVSSVAKRVLTMGVNGELHPSRFCEK	564
3NZ4_TcPAM	LDILKLMIAASHTAMCQAVDLRQLEELVKVVENVVSTLAD---ECGLPNDTKAR-----	527
1T6P_RtPAL	NDVLSLLLATHLYCVLQAIIDLRAIEFEFKQFGPAIVSLIDQHFGSAMTGSNLRD-----	571
2NYN_AvPAL	VDIFQNYVAIALMFGVQAVDLRKYKKTGHYDARAACLSPATER-----	510
2NYF_NpPAL	VDIFQNYMAIALMFGVQAVDLRKYKMGHYDARTCLSPNTVQ-----	510
	. : *	::
1GKM_PpHAL	-----SEVAHYDRDRFFAPDIEKAVELLAKGSLTGLLPAGVLP	507
2O6Y_RsTAL	-----EQFPPLETDRPLGQEI AALATHLLQQSPV-----	523
3UNV_PaPAM	-----TLVPYLEEDKTISDYIESIAQTV-----LTKNSDI	541
2OHY_SgTAM	-----RLVPTLGVDRYMADDIELVADALSRGEFLRAIARETDI	538
1W27_PcPAL	DLLRVVDREYIFAYIDDPCSATYPLMQKLRQTLVEHALKNGDNERNLSTSIFQKIATFED	624
3NZ4_TcPAM	-LLYVAKAVPVYTYLESPCDPTLPLLLGLKQSCFGSILALHKKDGIETDTLVDRLAEFEK	586
1T6P_RtPAL	-ELVEKVNKTLAKRLEQTNSYDLVPRWHDAFSFAAGTVVEVLSSTSLSLAAVNAAWKVAAA	630
2NYN_AvPAL	-----LYSAVRHVVGQKPTSDRPYIWNNDNEQGLDEHIARISADIAAGGVIVQAVQDILP	564
2NYF_NpPAL	-----LYTAVCEVVGKPLTSVRPYIWNNDNEQGLDEHIARISADIAGGLIVQAVEHIFS	564
	:	
1GKM_PpHAL	SL-----	509
2O6Y_RsTAL	-----	
3UNV_PaPAM	-----	
2OHY_SgTAM	QLR-----	541
1W27_PcPAL	ELKALLPKEVESARALES-GNPAIPN-----RIEECRSYPLYKFVRKELGT-----EYL	673
3NZ4_TcPAM	RLSDRLLENEMTAVRVLYEKKGHKTADNNDALVRIQGSRFLPFYRFVRDELDT-----GVM	641
1T6P_RtPAL	ESAI SLTRQVRETFWSAASTSSPALS Y-----LSPRTQILYAFVREELGVKARRGDVF	683
2NYN_AvPAL	CLH-----	567
2NYF_NpPAL	SLKST-----	569
1GKM_PpHAL	-----	
2O6Y_RsTAL	-----	
3UNV_PaPAM	-----	
2OHY_SgTAM	-----	
1W27_PcPAL	TGEKVTSPGEEFEKVFIAMSKGEIIDPLLECLESWNGAPLPIC-----	716
3NZ4_TcPAM	SARREQTPQEDVQKVFDAIADGRITVPLLHCLQGFLGQPNGCANGVESFQSVWNKSA	698
1T6P_RtPAL	LGKQEVITIGSNVSKIYEAIKSGRINNVLLKMLA-----	716
2NYN_AvPAL	-----	
2NYF_NpPAL	-----	

*Pp*HAL highlights the structural characteristics for this class (PDB ID 1GKM). The enzyme consists of mainly parallel alpha helices. The MIO marks the active site which rest at the end of a long four helix bundle. At either end of the monomer are large loop regions known as the inner and outer loops. Four of these monomers come together to form a tetramer of head to tail dimers. The active site, which is quadruply redundant within the tetramer, becomes encapsulated within the interface of three monomers and is enveloped by the inner and outer loops (22).

Based on the structure, the authors suggest a mechanism as follows (**Figure I.5**). First, histidine enters the active site where the C δ atom nucleophilically attacks the methylenide group of the MIO. This introduces a positive charge to the imidazole which acidifies the protons at the C β position. The H_{Re} proton from C β (as described in the literature) is abstracted by an enzymatic base, eliminating the α -ammonium group. Trans-urocanate is then released (22). Such a mechanism would be analogous to the Friedel-Crafts type reaction proposed for the aminomutases.

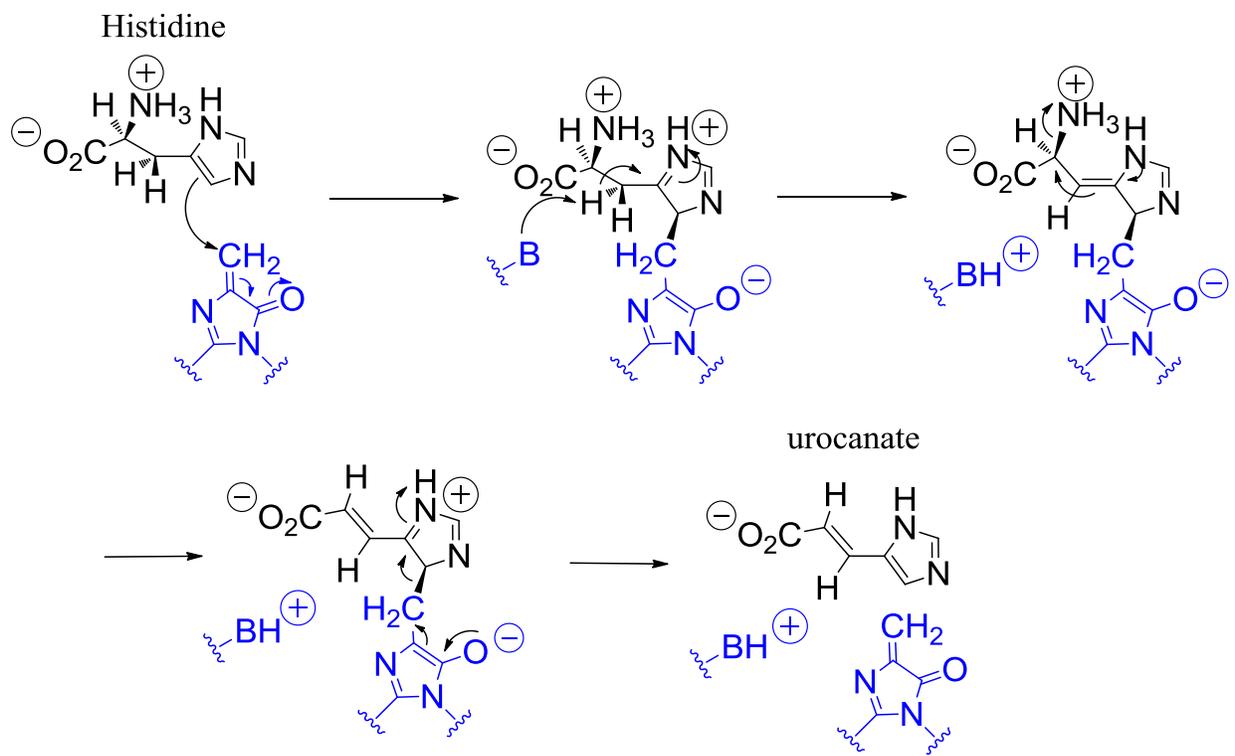


Figure I.5 The formation of urocanate in HALs as presented by Schwede et. al. (22) inspired the Friedel-Crafts type mechanism in PALs and PAMs.

PALs represent the first committed step in all phenylpropanoid biosynthesis, which would include lignins and a host of other natural products (33, 34) The first PAL structure was *Rhodospiridium toruloides* phenylalanine ammonia lyase (*Rt*PAL) from yeast which was solved in two different crystal forms (PDB IDs 1T6J and 1T6P) (21). It contained features similar to that of *Pp*HAL and other members of this family of enzymes with two notable exceptions. The first is the presence of a small capping domain at the end of the monomers which is common to plant PALs and PAMs (**Figure I.6**). The second exception was density off the MIO attributed to an NH₂ adduct, the observation of which was facilitated by the high crystallographic order of the active site which included electron density for the inner and outer loop regions (**Figure I.7**). The position of the MIO at the positive end of the four helix bundle inspired the authors to theorize that this arrangement would stabilize the charged intermediates found in the amino-group alkylation pathway by creating an electropositive MIO. Thus, the tertiary structure of the protein supports the amino-group alkylation pathway.

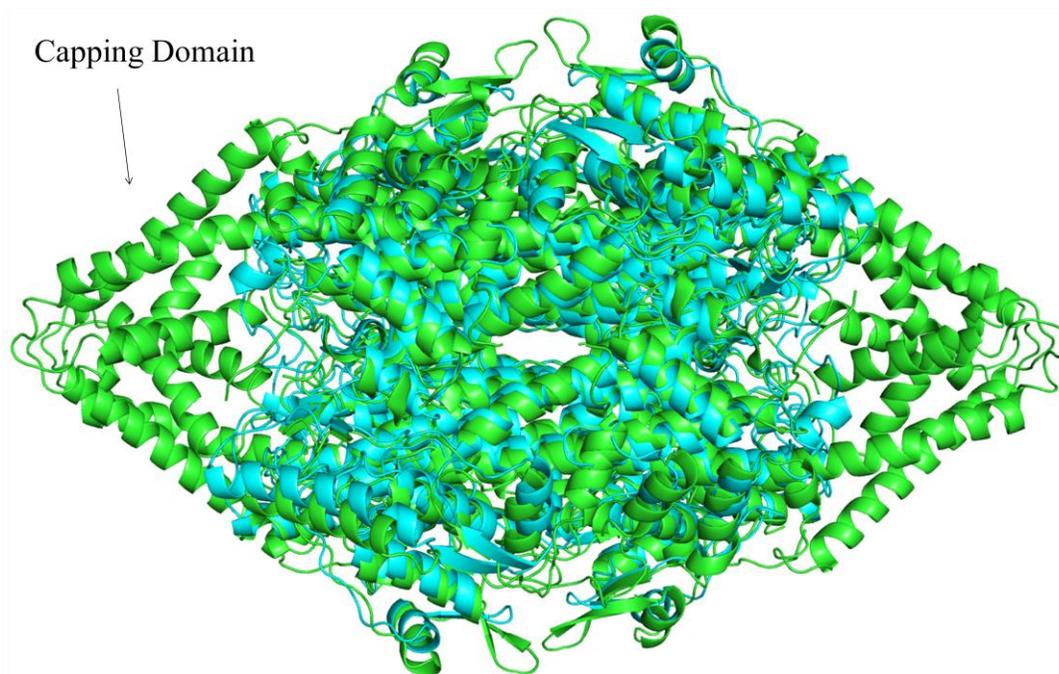


Figure I.6 Ribbon diagrams overlaid of RtPAL (green, PDB ID 1T6J) with PpHAL (blue, PDB ID 1GKM) highlight the capping domain present in plant PALs and PAMs.

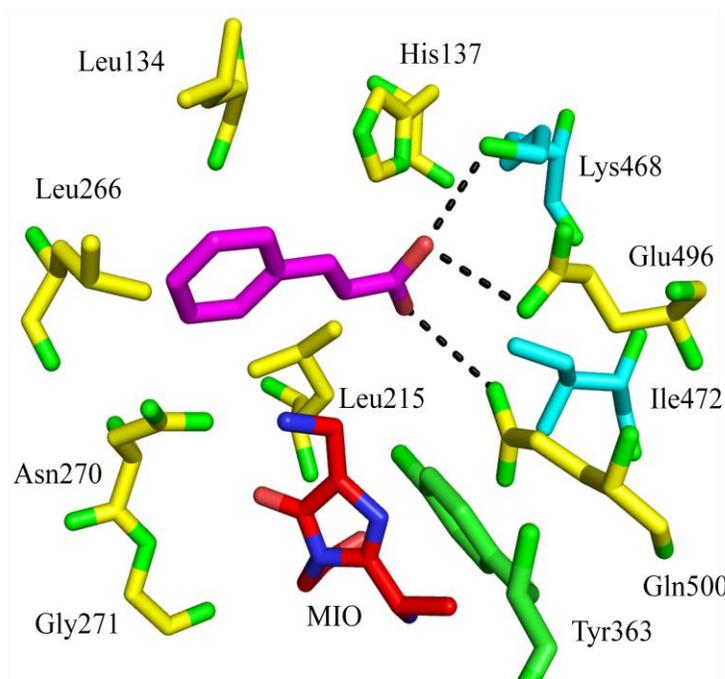


Figure I.7 Key residues in the active site of RtPAL are shown. Residues from monomers 1, 2 and 3 are shown in yellow, green and blue respectively. Cinnamate bound in the active site is shown in magenta. The MIO with an apparent ammonia adduct bound is highlighted in red. The trajectory of the cinnamate is atypical for this class of compounds.

PALs are also found in cyanobacteria such as *Anabaena Variabilis* ATCC 29413 (*AvPAL*) (PDB ID 2NYN) and *Nostoc punctiforme* ATCC 29133 (*NpPAL*) (PDB 2NYF). The structures show that the cyanobacterial PALs are similar to plant and yeast PALs as well as the HALs (5) with the exception that they lack a small N-terminal capping domain. This similarity to HALs suggests PALs may use a similar mechanistic pathway. Further studies into reducing the proteolytic susceptibility of the protein for its potential use as a treatment for the metabolic disorder phenylketonuria lead to a structure of Cys503Ser/Cys565Ser double mutant *AvPAL* which contained the first well ordered active site loops along with density for a cinnamic acid (PDB ID 3CZO). Docking studies of this structure suggested the enzyme uses the amino-group alkylation pathway (35).

Similar automated docking and molecular dynamics (MD) simulation studies were performed on the structure of *Petroselinum crispum* phenylalanine ammonia lyase (*PcPAL*) (PDB ID 1W27) to determine if it followed the amino-group alkylation pathway or the Friedel-Craft type mechanism (36, 37). It has been reported that this structure represents an inactive form as the catalytic Tyr110 (Tyr60 in *RsTAL*) is far from the active site. MD simulations to place the Tyr110 containing loop into the correct orientation highlighted the possible importance of Glu484, which was shown to interact with the substrate amino group pulling it away from the MIO. Such an orientation would support the Friedel-Crafts mechanism in PALs and PAMs but an amino-group alkylation pathway in TALs and TAMs (38).

Wild-type *Rhodobacter sphaeroides* TAL (*RsTAL*) has been determined both unligated (PDB ID 2O6Y) and with several different ligands bound to the active site including coumarate (PDB ID 2O7B) and caffeate (PDB ID 2O7D). The His89Phe mutant of *RsTAL* was determined with the inhibitor 2-aminoindan-2-phosphonic acid (AIP) bound (PDB ID 2O7E), cinnamate bound (PDB ID 2O78) and coumarate bound (PDB ID 2O7F). The structures demonstrate the highly conserved Tyr60 and Arg303 interact with the propenoate moiety as might be expected for an amino-group alkylation pathway. The structure of the AIP bound to His89Phe mutant *RsTAL* shows the attachment of the MIO at the methylidene group to the amino group of the acid and the ability of the TAL to switch selectivity to a PAL with a single point mutation (39).

SgcC4 L-tyrosine 2,3-aminomutase (*SgTAM*) represents the first tyrosine aminomutase whose structure was determined (40). Sequence similarities between TAMs and TALs suggest the mechanism of action is likely related. The structure of *SgTAM* (PDB ID 2OHY) with the bound inhibitor α,α -difluoro- β -tyrosine (PDB ID 2QVE), inhibitor (2S,3S)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid (PDB ID 2RJR) and alternatively the substrate mimic (3R)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid has been determined (PDB ID 2RJS) (41) represent the first structures of a TAM. The attachment of the inhibitors to the MIO is suggestive of an amino-group alkylation pathway (**Figure I.8**). The crystal structure of the Tyr63Phe mutant *SgTAM* with bound tyrosine (PDB ID 3KDZ) shows the important catalytic function of the highly conserved Tyr63 (Tyr110 in *PcPAL* and Tyr60 in *RsTAL*) as it relates to the amino-group alkylation pathway (42). The tight packing of the active site is proposed to keep water out of the binding pocket which allows for the retention of ammonia, enforcing aminomutases rather than ammonia lyase activity.

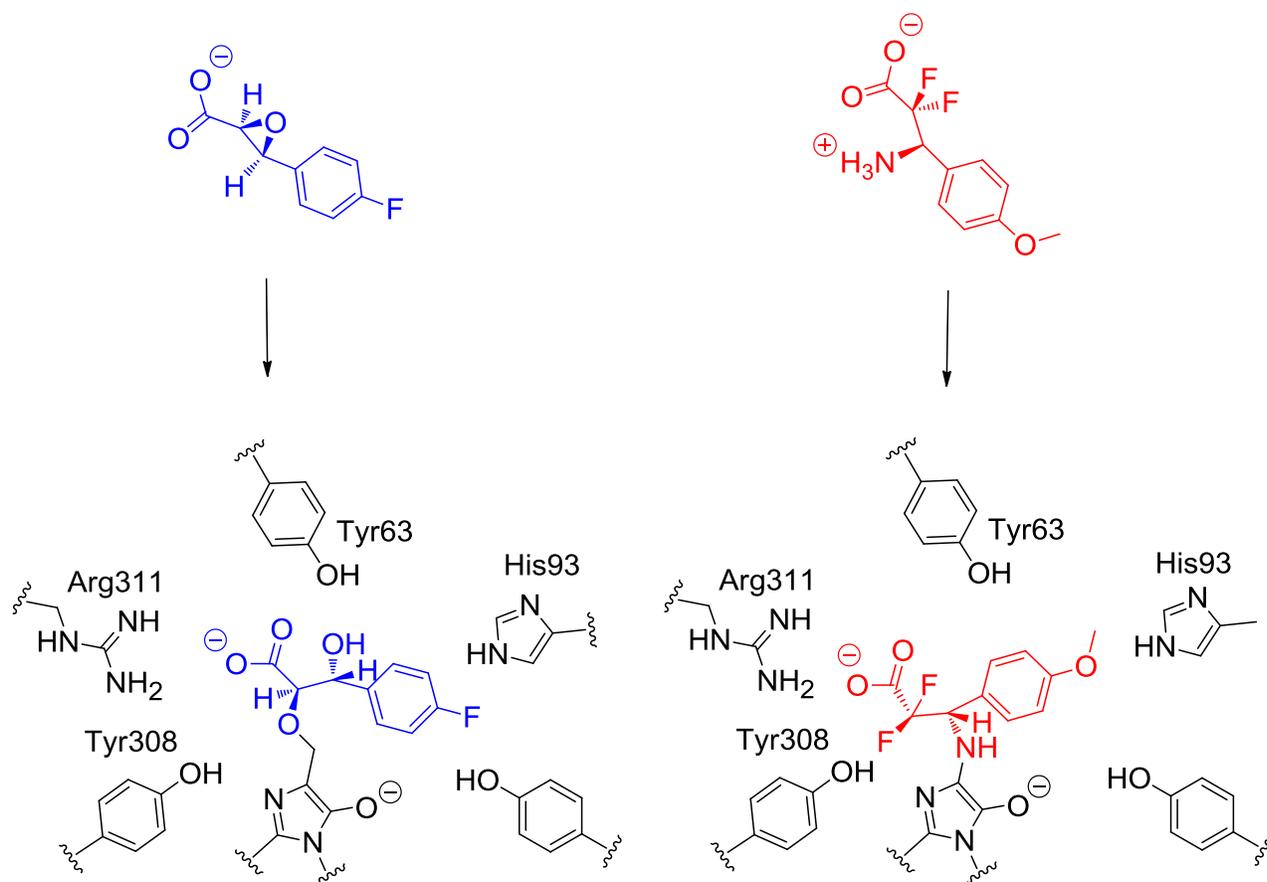


Figure I.8 Active site of *SgTAM* with bound inhibitor (2*S*,3*S*)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid (blue) (left, PDB ID 2RJR) and the substrate mimic (3*R*)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (red) (right, PDB ID 2RJS) suggests an amino-group alkylation pathway.

The crystal structure of *Taxus canadensis* phenylalanine aminomutase (*TcPAM*) has recently been determined with cinnamate bound to the active site (43). As stated above, *TcPAM* converts (*S*)- α -phenylalanine to (*R*)- β -phenylalanine. Like other members of this family, the protein exists as a tetramer with the interface of three monomers being responsible for the formation of the active sites. The key residues within the active site which are shown in **Figure I.9** are similar to those found in PaPAM, suggesting similar mechanistic pathways.

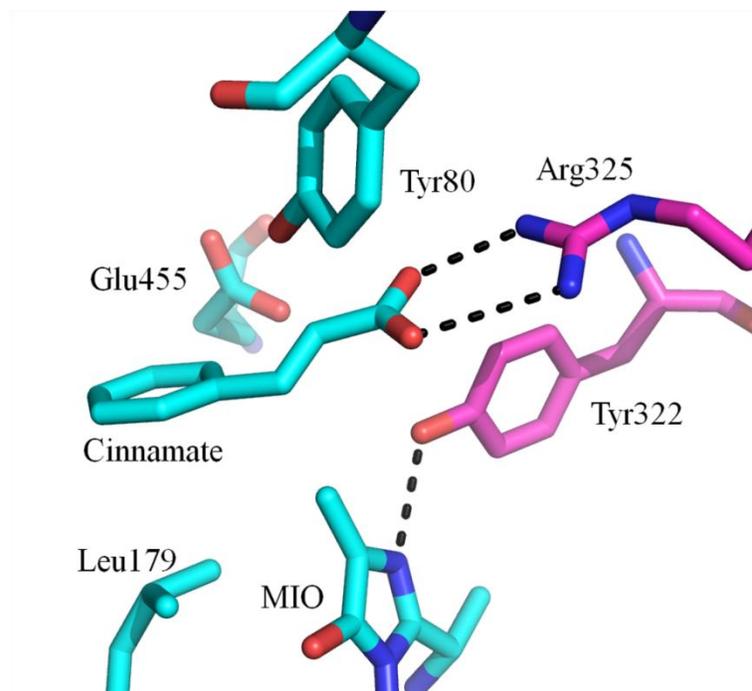


Figure I.9 The key TcPAM residues that interact with cinnamate. Residues of “subunit A” are colored blue and residues of “subunit B” are colored magenta.

The residues on TcPAM responsible for turnover of the enzyme were identified via mutagenesis (43). Tyr80 is positioned to remove the benzylic hydrogen from the bound phenylalanine as would be necessary for the amino-group alkylation pathway. This same residue is present in other structures (Tyr63 in SgTAM, Tyr110 in *PcPAL*, and Tyr60 in *RsTAL*) where it is believed to contribute in a similar fashion within the amino-group alkylation pathway (29 and references therein).

Three monomers combine to form the active site. Considering the “subunit A” to contain the catalytic Tyr80, Tyr322 of the “subunit B” also contributes to the activity of the enzyme, presumably by stabilizing the polar MIO intermediates that are formed as the substrate interacts with the MIO. Also critical is Arg325 of the “subunit B” which forms a double handed salt bridge to the carboxylate of the cinnamate, holding it in place. This positions the MIO to either accept the amino group or react with the aromatic ring, forming the cinnamate intermediate shown. Residues Leu104, Leu179, Leu227 and Val 230 (all of “subunit A”) and Lys 427 and Ile431 (of “subunit C”) form a hydrophobic pocket for the phenylalanine ring to reside in.

However, the current state of the structure does not imply either mechanism is correct since the formation of cinnamate is mutual to both schemes. The presence of Glu455 is suggestive of a Friedel-Crafts-type mechanism (31). However, the carboxylate is firmly positioned by Arg325 with Tyr80 positioned for reaction with the propenoate moiety as would be expected for the amino-group alkylation pathway. It has been shown that the benzylic hydrogen atom and amino group trade positions with facial selectivity in *TcPAM* as expected for the amino-group alkylation pathway (44).

The gram-negative bacteria *Pantoea agglomerans* is an opportunistic pathogen that is often found in the stomachs of locust. It produces a PAM (*Pa*PAM) that represents an interesting target for crystallization because *Pa*PAM (which is also referred to as AdmH) is responsible for the formation of (3S)- β -phenylalanine from naturally occurring (3S)- α -phenylalanine in the antibiotic Andrimid biosynthetic pathway (4). The (3S)- β -phenylalanine product is adenylated by AdmJ for conversion to the final product, Andrimid (45). The initial transformation is similar to that of *Tc*PAM and *Sg*TAM, making comparisons between the three possible. Much information is known about the promiscuity of this enzyme and is soon to be published. By obtaining the crystal structure we might gain insight into the mechanism utilized by this family of enzymes.

I.2. Experimental Procedures

I.2.1. Crystallization of *Pantoea agglomerans* Phenylalanine Aminomutase

Protein for crystallization was provided by Udayanga Wanninayake of the Walker Lab. Twelve liters of *Escherichia coli* cell cultures were grown in LB medium supplemented with kanamycin (50 $\mu\text{g}/\text{mL}$), induced with 100 μM IPTG at 16°C for 16 hours before harvesting by centrifugation. The cell pellet was dissolved in 125 mL of resuspension buffer (50mM sodium phosphate (pH 8.0), 5% (v/v) glycerol, 300mM NaCl and 10mM imidazole), lysed by sonication, and supernatant clarified by centrifugation. The His-tagged protein was purified by Ni NTA affinity chromatography (Qiagen) with the protein eluting in 250 mM imidazole containing buffer before being buffer exchanged into 50 mM sodium phosphate buffer with 5% glycerol and concentrated to 7.0mg/mL using Centriprep centrifugal filters. SDS-Page analysis of the protein showed it to be >95% pure. Cinnamate in a ration of 10:1 ligand: protein was added.

The Gryphon LCP robot (Art Robbins) was used to screen six different crystallization screens of 96 conditions a piece using the sitting drop method with plates set both at room temperature and 4°C. Approximately 13 different conditions at room temperature produced crystals overnight of varying quality; the best being from the Salt RX Screen (Hampton Research) condition F12 (1.5 M Lithium sulfate monohydrate, 0.1M Tris-HCl, pH 8.5). Screening of this condition using the hanging-drop diffusion method to 0.1M Tris-HCl, pH 7.5, and 2.5M lithium sulfate produced crystals appropriate for x-ray data collection. Crystals of PaPAM appeared after one day, growing to full size in two days of dimensions 1mm X 0.5mm X 0.5mm (**Figure I.10**).

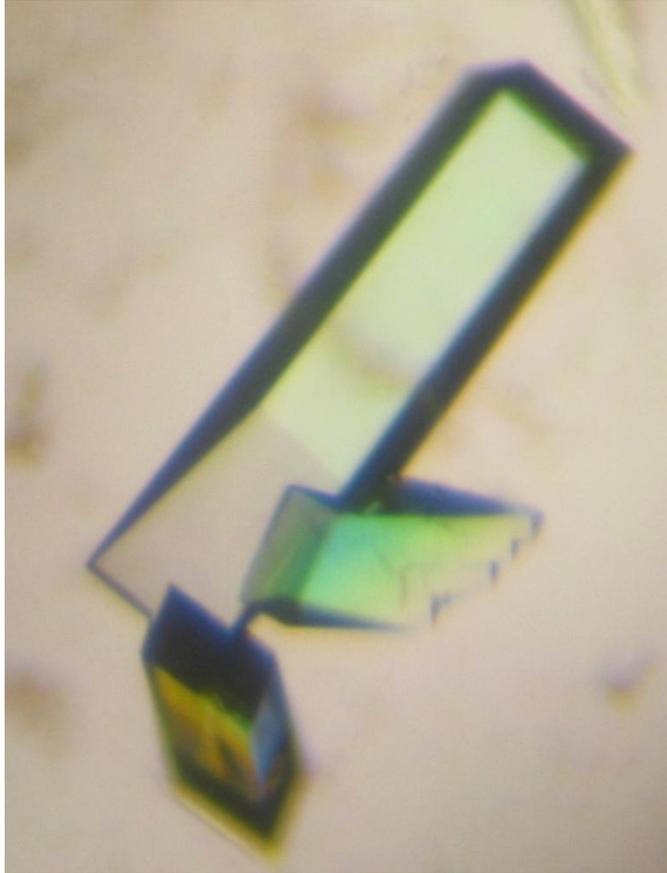


Figure I.10 Crystals of *Pantoea agglomerans* Phenylalanine Aminomutase.

Crystals were soaked in cryoprotectant (0.1M Tris-HCl, pH 7.5, 2.5M lithium sulfate, 15% glycerol), mounted in CrypLoops (Hampton Research) and flash frozen in liquid nitrogen.

I.2.2. Structure Determination

Data was collected at the LS-CAT beamline at Argonne National Laboratories. 180 1° images were collected, but only the first 100 were ultimately used due to decay. Raw diffraction data was indexed, processed and scaled using the HKL 2000 package (56). A search was performed using the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST). This indicated *Streptomyces globisporus* tyrosine aminomutase (*SgTAM*) (PDB ID 3KDY) to be the closest homolog. Swiss-Model (46) produced a threaded homology model based on the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) model of 3KDY and the known amino acid sequence of PAM. The structure was solved using this model and the MOLREP program in the CCP4 suite of programs (47). Refmac in the CCP4 suite produced the initial electron density maps, which were further corrected manually in COOT (**Table I.2**). All refinements were performed using REFMAC. Jligand version 1.0.9 (48) was used to generate the necessary ligands.

Table I.2 Data-Collection and Structure-Refinement Statistics for *PaPAM*

	data collection
wavelength (Å)	0.97872
total reflections	385668
unique reflections	1307130
space group	C 2 2 2
unit-cell parameters	$a = 153.96, b = 185.84, c = 72.56$ $\alpha = \gamma = \beta = 90^\circ$
molecules per ASU	2
resolution range (Å)	50-1.7 (1.73-1.70)
completeness (%)	93.6 (71.9)
I/ σ	13.1 (2.5)
R _{merge} (%)	0.11 (0.86)
	structure refinement
resolution (Å)	1.70
R _{cryst} /R _{free} (%)	0.167/0.223
	rmsd from ideal values
bond length (Å)	0.024
bond angle (deg)	1.958
average <i>B</i> factor	21.712
PDB #	3UNV

I.3. Results and Discussion

I.3.1. Crystal Structure of *Pantoea agglomerans* Phenylalanine Aminomutase

PaPAM is similar in architecture to other bacterial lyases. It lacks the C-terminal capping domain found in the related plant lyases such as *TcPAM* (**Figure I.11**). Present is a well defined outer loop region which interacts with adjacent monomers to form the active tetramer. The inner loop region which rests just above the active site is packed tighter towards the MIO than that of *TcPAM* and is well ordered (**Figure I.12**).

PaPAM consists of mostly α -helices that run parallel to one another, forming at the center a four helix bundle. At the end of this bundle is the active site, which exists at the monomer face (**Figure I.13**). The *PaPAM* MIO moiety differs in arrangement from other lyases with the sequence Thr-Ser-Gly auto-catalytically forming the MIO as evident in the electron density map. Three monomers combine to define the MIO containing active site which is quadruply redundant in the tetramer (**Figure I.14**). The active site was found to contain (3S)- β -phenylalanine attached via its amino group to the methyldiene of the MIO in the case of the arbitrarily assigned “C” and “D subunits” while the “A” and “B subunits” contained a mixture of (2S)- α - and (3S)- β -phenylalanine attached via their amino groups to the methyldiene of the MIO (**Figure I.15**).

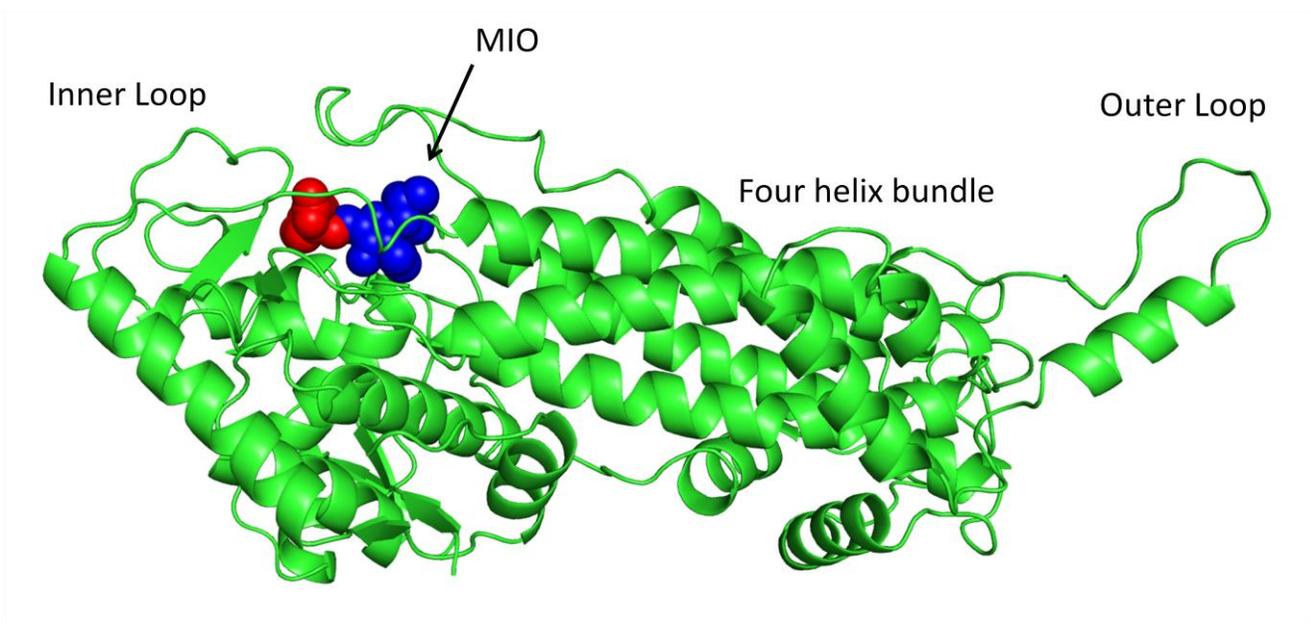


Figure I.11 *PaPAM* monomer (green) showing the inner and outer loop regions (as marked), MIO (blue spheres) and (3*S*)- β -phenylalanine (red spheres). The active site rests at the monomer's surface at the end of a long four helix bundle.

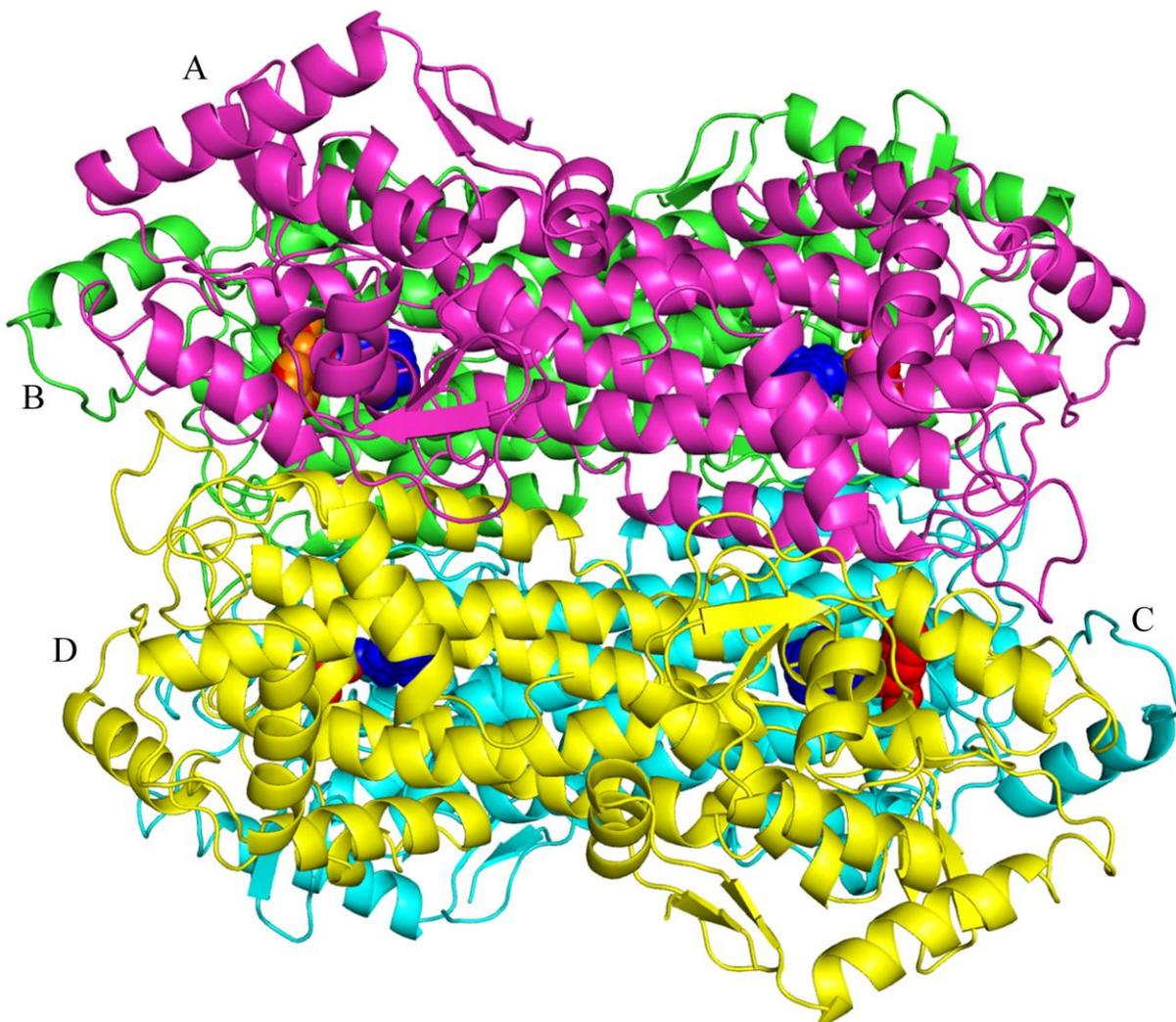


Figure I.12 The PaPAM tetramer showing the relationship between the four monomers is that of a head-to-tail arrangement. Four active sites rest at the interfaces created by the tetramer. The residues from three monomers contribute to the active sites including residues from the inner loop of the MIO containing monomer to the outer loop residues of an adjacent monomer. The MIO is highlighted in blue spheres. The (3S)- β -phenylalanine is highlighted in red spheres and (2S)- α -phenylalanine is highlighted in orange spheres which are seen overlapping with those of the (3S)- β -phenylalanine.

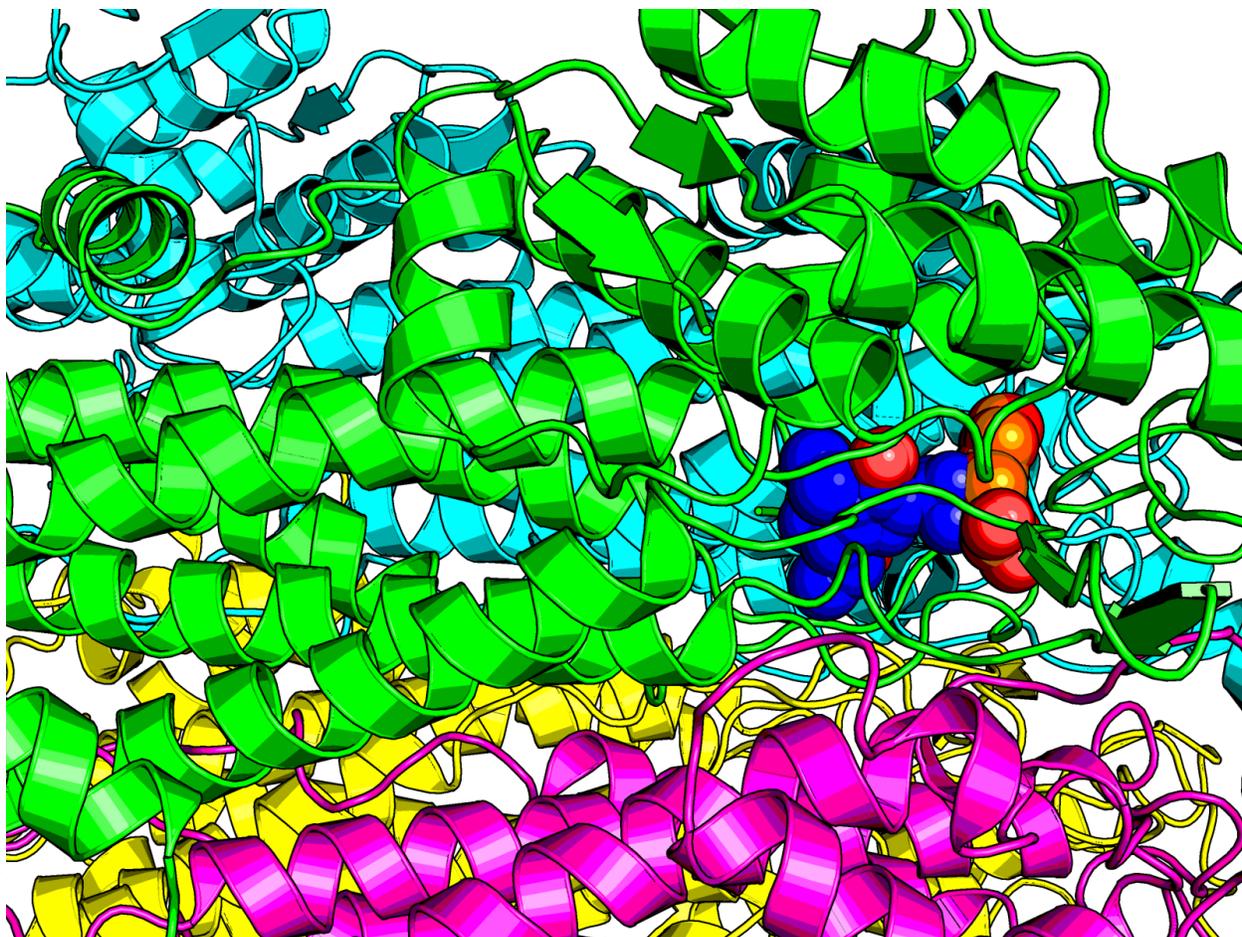


Figure I.13 Each of the four MIOs (blue spheres) within the tetrameric form of PaPAM is positioned at the end of a long four-helix bundle (see green helixes on left hand side of image) which are believed to stabilize negative charges generated on the MIO during the amino-group alkylation pathway. Ribbon diagrams of the four separate monomers are shown in magenta, green, yellow and aqua. The (3S)- β -phenylalanine is highlighted in red spheres and (3S)- α -phenylalanine in highlighted in orange spheres.

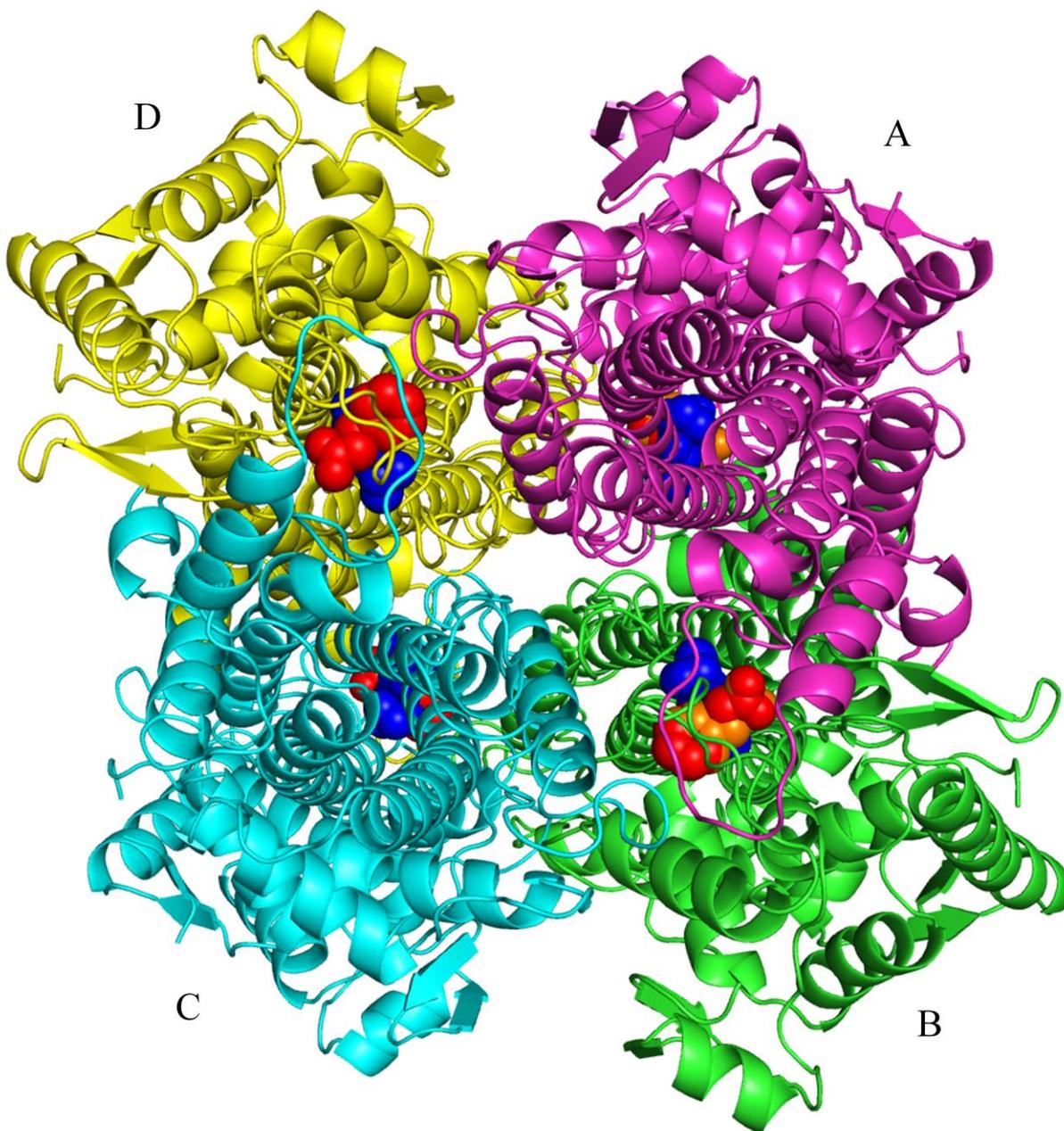


Figure I.14 The PaPAM tetramer showing the relationship between the four monomers and the interactions between the inner and outer loop regions and the MIO active site. The inner loop of the monomer containing the MIO is packed between the active site and the outer loop region of a neighboring monomer. The MIO is highlighted in blue spheres. The (3S)-β-phenylalanine is highlighted in red spheres and (3S)-α-phenylalanine in highlighted in orange spheres.

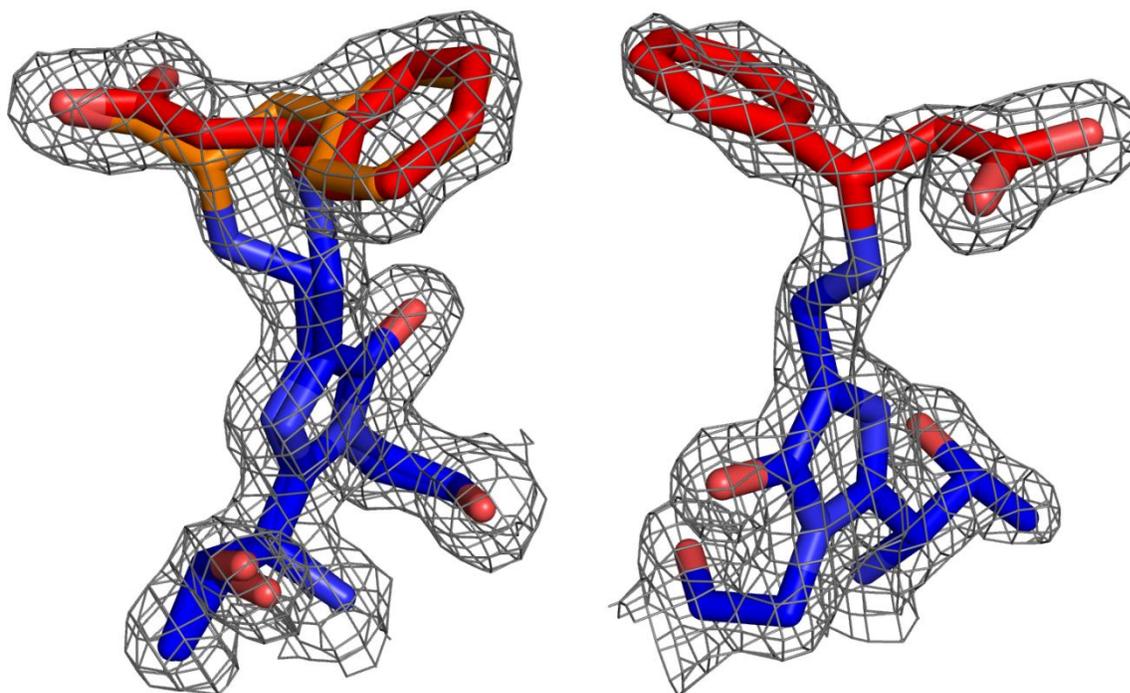


Figure I.15 The active site of chain A in *PaPAM* (left) showing both (2S)- α -phenylalanine (orange) and (3S)- β -phenylalanine (red) attached to the MIO (blue) and the active site of chain B (right) showing (3S)- β -phenylalanine alone attached to the MIO. Such an arrangement of the ligands is a clear indication of an amino-group alkylation pathway. The grey mesh represents the electron density around the ligands observed at 1.2 sigma.

The *PaPAM* active site contains residues that are highly conserved among this family of enzymes (**Figure I.16**). Catalytically significant Tyr78 and Tyr320 (Tyr80 and Tyr322 in *TcPAM*, Tyr63 and Tyr308 in *SgTAM*, and Tyr60 and Tyr300 in *RsTAL*) are positioned facing the α -hydrogens of the bound acid. The close proximity of Tyr78 to the MIO mimics Tyr63 found in *SgTAM*, where Tyr80 of *TcPAM* is brought further out of the active site by the inner loop region. The hydrophobic Val108 (Leu108 in *TcPAM*), corresponds to His93 in *SgTAM* and His89 in *RsTAL*, which lends their active sites for tyrosine binding. Residues Leu171, Leu216 and Ile219 further define the hydrophobic binding pocket of the phenyl ring. Unlike *TcPAM* and *PcPAL* which have Glu455 and Glu484 respectively or *SgTAM* and *RsTAL* which have Asn438 and Asn432 respectively, *PaPAM* has Thr452 close to the MIO.

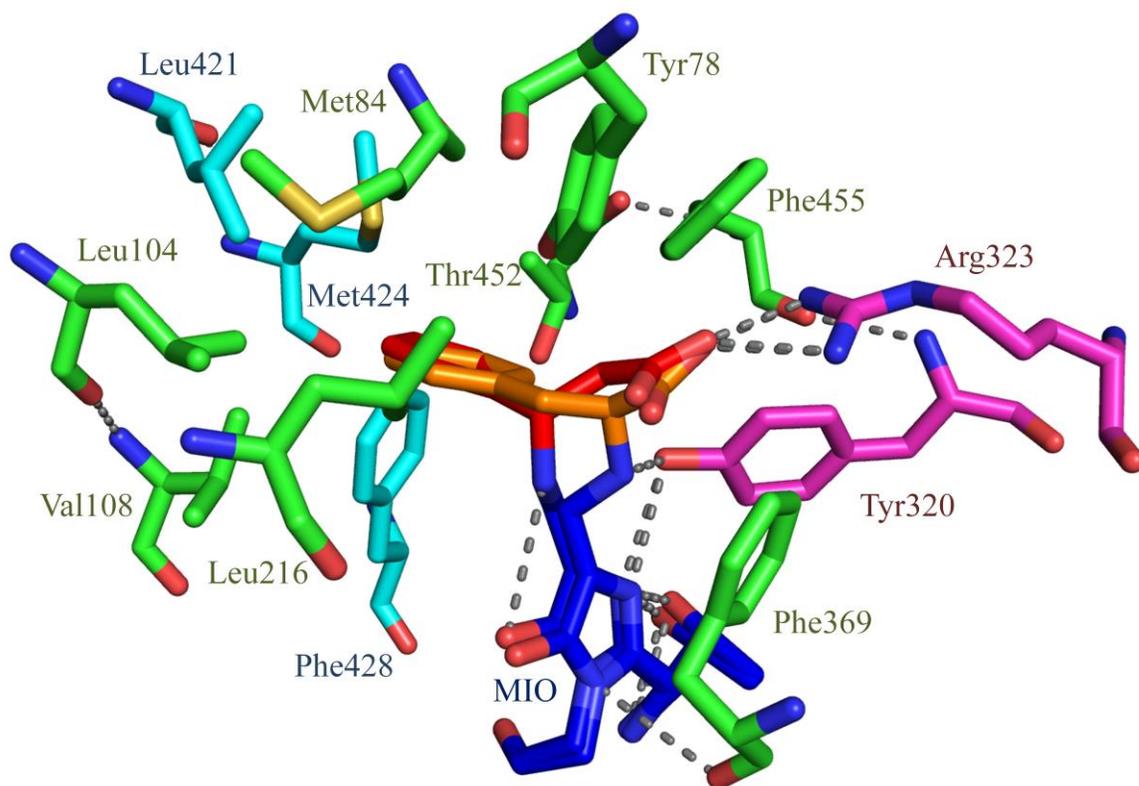


Figure I.16 Key residues in the MIO active site of *PaPAM* “subunit B” are highlighted in green. “Subunit A” is highlighted in magenta and “subunit C” is highlighted in cyan. The MIO is highlighted in blue. The (3*S*)- β -phenylalanine is highlighted in red and (2*S*)- α -phenylalanine in highlighted in orange. Residues from all three subunits are needed to form the active site.

When compared to the active site of *RtPAL* we see similarities in the known catalytic residues such as Tyr320 (Tyr363 in *RtPAL*) and Arg323 (Arg366 in *RtPAL*). Where Tyr320 in *PaPAM* occupies a position similar to that of Tyr363 in *RtPAL*, Arg323 in *PaPAM* is positioned towards the carboxylates of the ligands attached to the MIO. As no ligands are present in *RtPAL*, Arg366 is instead twisted somewhat away from the active site of *RtPAL* (**Figure I.17**).

This suggests that Arg366 is flexible and that hydrogen bonding interactions with the ligand help to hold both Arg366 and the ligand in place. This flexibility is likely also present in Arg 323 of *PaPAM*. The ammonia adduct on the MIO of *RtPAL* is positioned directly between the two phenylalanine constitutional isomers. It demonstrates the possible trajectory that the ammonia group would travel in PAMs as the reaction progresses from the α to the β form of the product.

Lastly, we note that Tyr78 in *PaPAM* overlaps with the phenyl ring of cinnamate in *RtPAL*. The position of the cinnamate is not consistent with the positions of the phenylalanine ligands within the *PaPAM* active site, thus its position does not likely represent the location phenylalanine or cinnamate would occupy during the reaction the cinnamate in *RtPAL*. Rather, it appears the cinnamate is positioned within the same pocket that the catalytic Tyr110 would occupy when the inner loop of *RtPAL* presses into the active site. This loop is not visible in the structure of *RtPAL* due to its flexibility.

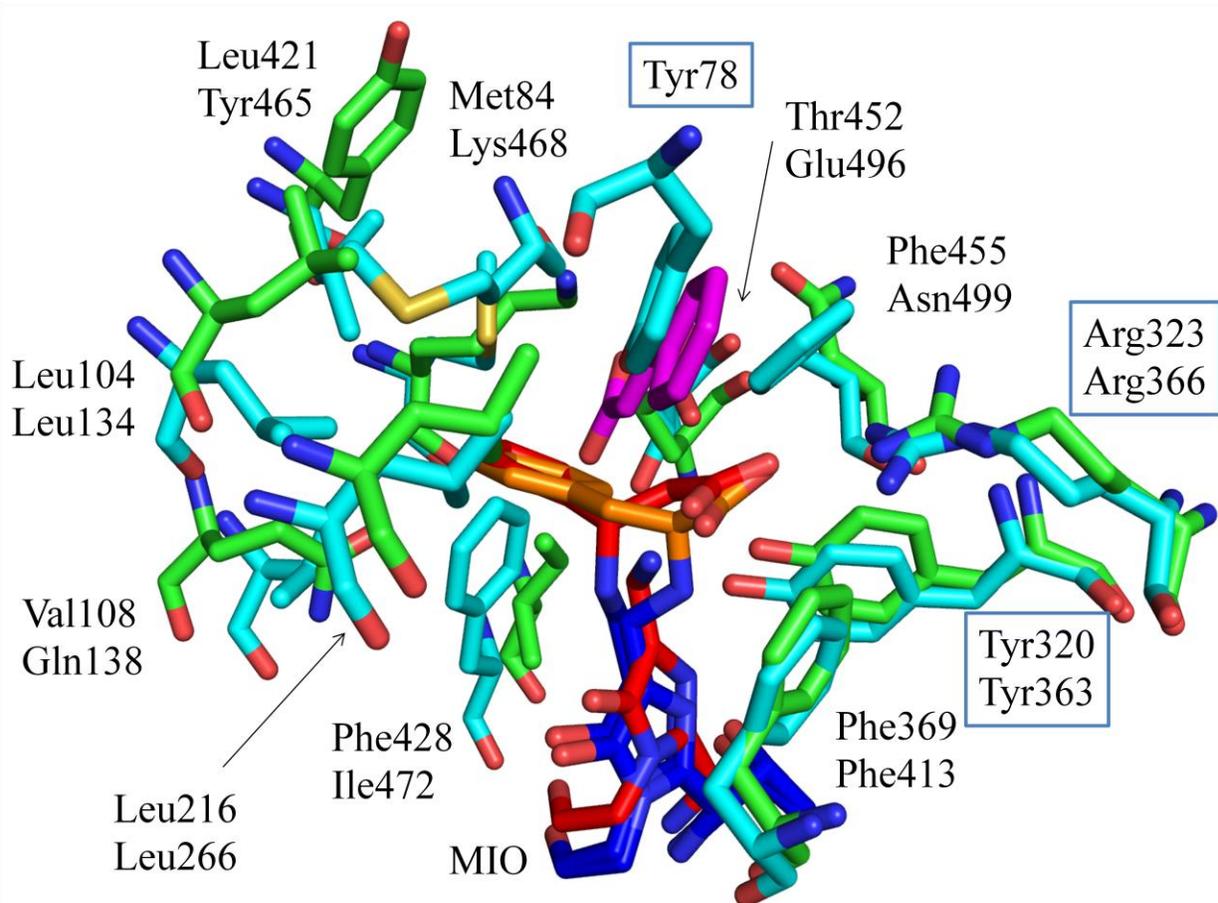


Figure I.17 The active site residues of *PaPAM* (cyan, PDB ID 3UNV) are overlaid with those of *RtPAL* (green, PDB ID 1T6J). The left hand side shows the various hydrophobic residues that make up the binding site for the phenyl ring of the phenylalanine ligand with residue identification numbers for *PaPAM* appearing on top of those for *RtPAL*. Certain conserved residues are highlighted in blue boxes. The ammonia adduct attached to the methylidene of the *RtPAL* MIO (red) is positioned between the two nitrogens of the phenylalanine ligands of *PaPAM*: (2*S*)- α -phenylalanine (orange) and (3*S*)- β -phenylalanine (red) which are attached to the MIO of *PaPAM* (blue). Cinnamate belonging to the *RtPAL* structure is highlighted in magenta and rests just above the active site in line with Tyr78 of *PaPAM*.

In *Pa*PAM this loop is highly ordered suggesting it is more rigid. Therefore the flexibility of the inner and outer loops may be responsible for the PAM versus PAL activity in this class of enzymes. Flexibility in the loops would allow for the cinnamic acid to be easily released from the active site before the ammonia has a chance to react to the β position. This movement opens up a surface binding site for the cinnamate that was previously occupied by Tyr110, pulling the cinnamate away from the MIO. Contrarily, in *Pa*PAM the loop is packed tightly against the active site, holding the cinnamate in place for further reaction.

Comparison of *Pa*PAM with *Sg*TAM shows that the preference for tyrosine over phenylalanine is due to polar residues located near the para position of the phenyl rings of either ligand (**Figure I.18**). Compared to *Rt*PAL which mimicked the hydrophobic residues found in *Pa*PAM, there are many charged residues which help to position the hydroxyl group of the tyrosine.

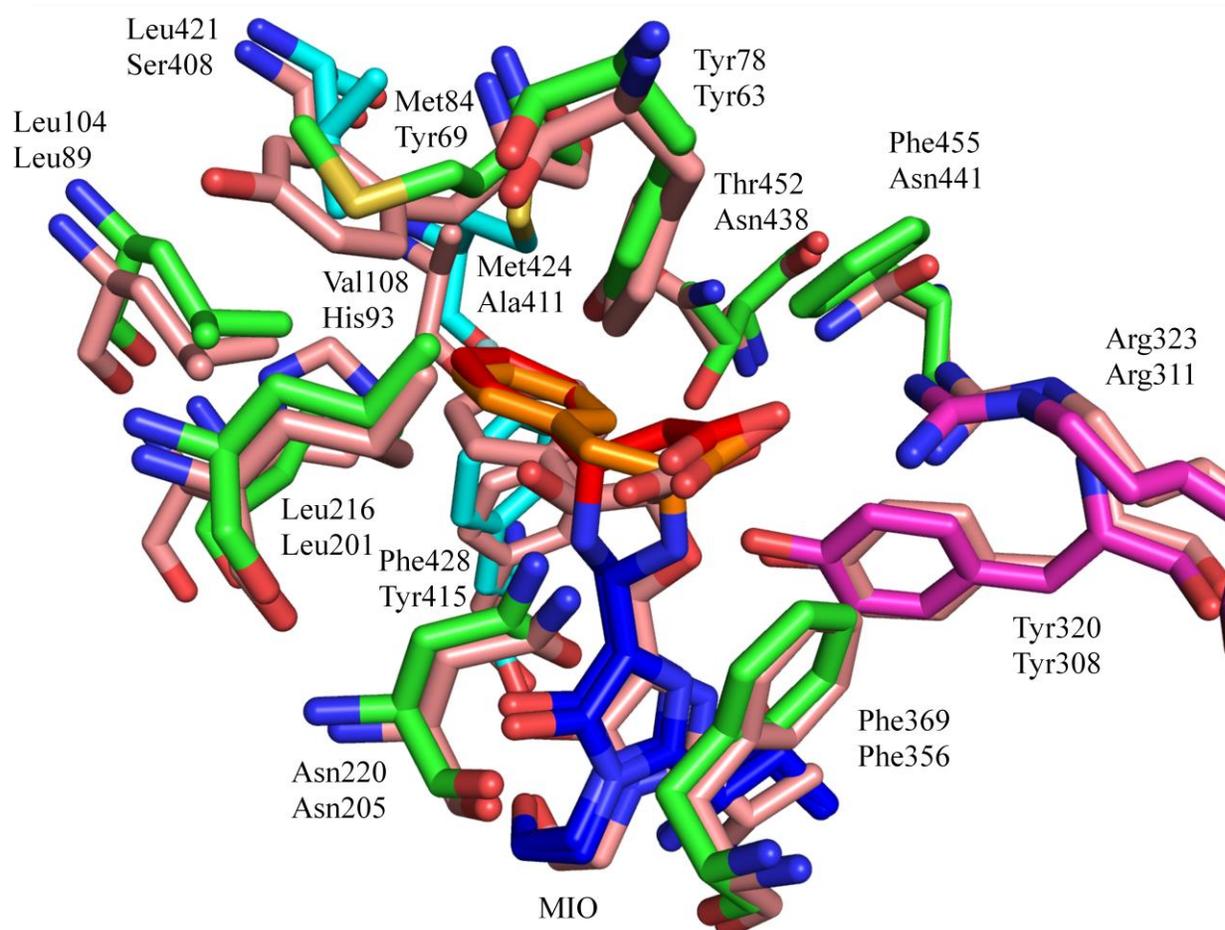


Figure I.18 Overlay of *PaPAM* residues for the various monomers (shown in green, magenta and cyan) with bound ligands (2*S*)- α -phenylalanine (orange) and (3*S*)- β -phenylalanine (red) which are attached to the MIO (blue) (PDB ID 3UNV) and *SgTAM* with bound inhibitor (2*S*,3*S*)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid (pink, PDB ID 2RJR) shows many charged residues around the phenyl ring in *SgTAM* are non-polar in *PaPAM* giving rise to the difference in substrate specificity. In addition to the ligands, the MIOs, Arg323/Arg311, Asn220/Asn205, Phe369/Phe356 and Tyr320/Tyr308 overlap one another, suggesting the control of substrate specificity is the same for both enzymes.

Despite the differences in ligands, both *Pa*PAM and *Sg*TAM give product with the same stereochemistry. Comparisons of the active site residues of *Sg*TAM with the inhibitor (2*S*,3*S*)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid (PDB ID 2RJR) and the substrate mimic (3*R*)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (PDB ID 2RJS) show the residues in the active site are in essentially identical positions for both structures (**Figure I.19**). This suggests that there are no large movements of residues during the turnover of the ligands. When these same residues are compared to *Pa*PAM we again see several residues with almost identical positions such as Arg323 (Arg311 in *Sg*TAM), Tyr78 (Tyr63 in *Sg*TAM) and Tyr320 (Tyr308 in *Sg*TAM). These three residues are known to be important in positioning the ligand for reaction (Arg323) or for the reactivity of the enzyme (Tyr78 and Tyr320). Residues with identical placement also include Phe369 (Phe356 in *Sg*TAM) and Asn220 (Asn205 in *Sg*TAM). These residues are close to the MIO and therefore might be important in enforcing the known stereochemistry of the reaction.

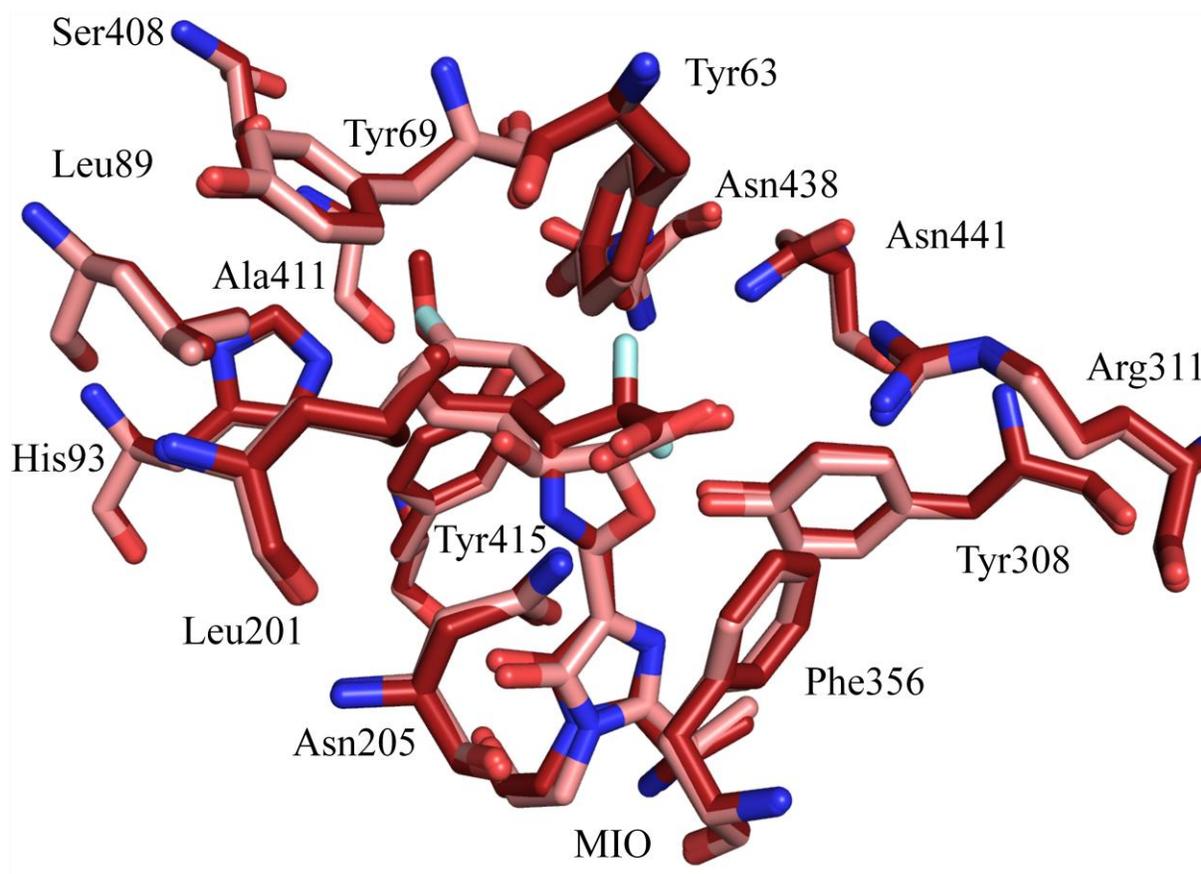


Figure I.19 Overlay of the active site residues of SgTAM with the inhibitor (2S, 3S)-3-(4-fluorophenyl)-2, 3-dihydroxypropanoic acid (pink, PDB ID 2RJR) and the substrate mimic (3R)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (red, PDB ID 2RJS) show the residues in the active site are in essentially identical positions for both structures. This suggests that movement of residues is not necessary for reaction and therefore residues occupying the same positions in *Pa*PAM are aiding in the stereoselectivity of both enzymes.

As the positions of the residues within the active site of SgTAM do not change significantly between the inhibitor (2S,3S)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid bound structure (PDB ID 2RJR) and the substrate mimic (3R)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid bound structure (PDB ID 2RJS) comparisons of the positions of these residues to *Pa*PAM in the substrate mimic (3R)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid bound structure hold similar conclusions to those already stated for the inhibitor (2S,3S)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid bound structure (PDB ID 2RJR). The key difference is in the location of the substrate mimic in SgTAM, the position of which matches well with that of the (3S)- β -phenylalanine (**Figure I.20**).

From the side view it becomes clear that compared to the positions of the un-substituted phenyl rings in *Pa*PAM, in the substrate mimic bound SgTAM structure the phenyl ring is somewhat pushed towards the MIO due to the methoxy group at the para position of the phenyl ring. This highlights the tight packed nature of the active sites, again suggesting that the flexibility of the inner and outer loops contribute to aminomutase versus lyase activity as the active sites of *Pa*PAM and SgTAM are tightly packed.

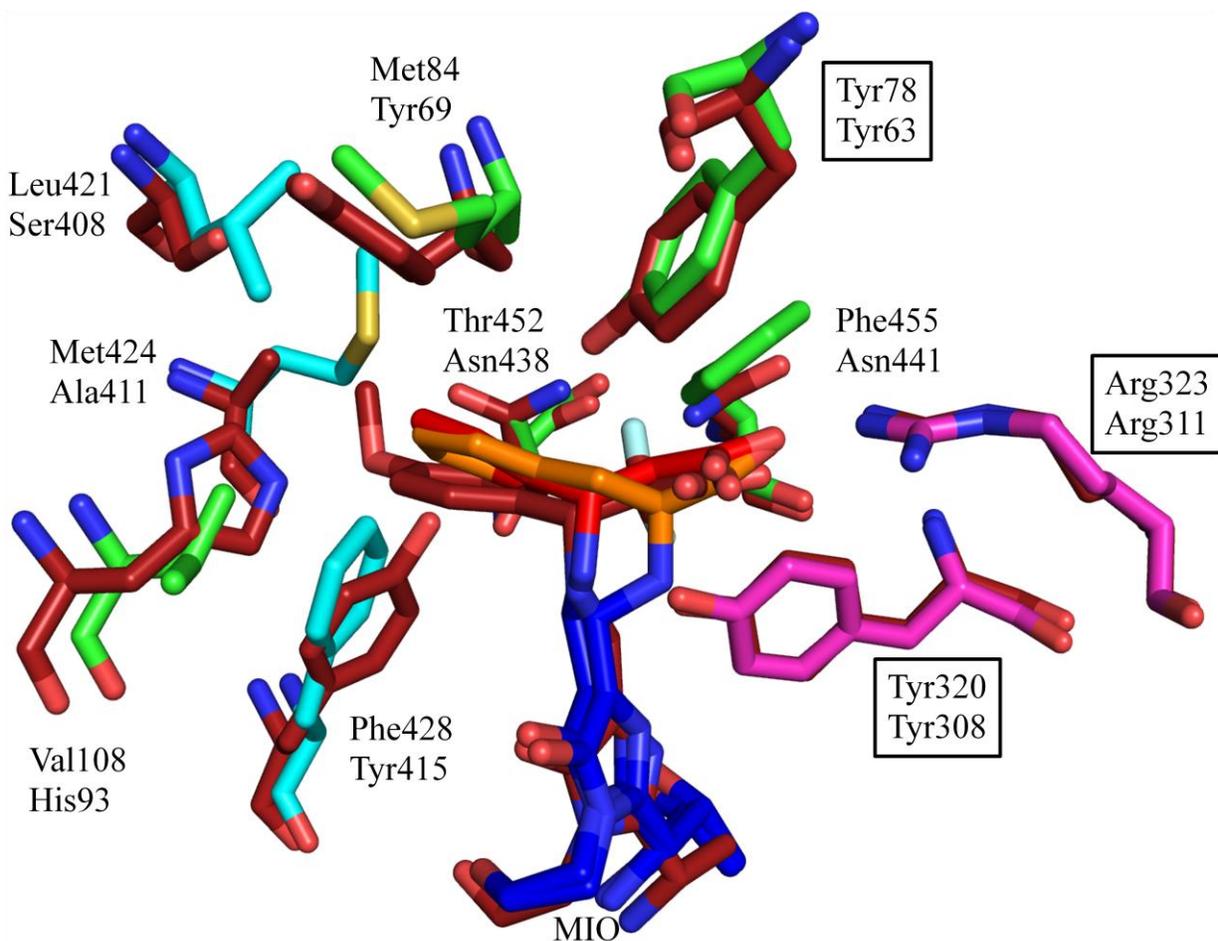


Figure I.20 Overlay of *PaPAM* residues for the various monomers (shown in green, magenta and cyan) with bound ligands (2S)- α -phenylalanine (orange) and (3S)- β -phenylalanine (red) which are attached to the MIO (blue) (PDB ID 3UNV) and *SgTAM* with bound substrate mimic (3R)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (red, PDB ID 2RJS) shows the depression of the para substituted phenyl ring of the substrate bound *SgTAM* compared to the location of the un-substituted phenyl rings of *PaPAM*. Such a movement towards the MIO suggests tight packing interactions of the inner and outer loops, which likely control the preference for aminomutase activity over lyase activity in *PaPAM* and *SgTAM*.

The amino group of the (3S)- β -phenylalanine is positioned towards the keto group oxygen of the MIO as found in *SgTAM* (23) *SgTAM* with bound substrate mimic (3R)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid. Likewise the (2S)- α -phenylalanine in *PaPAM* copies the trajectory of the inhibitor (2S, 3S)-3-(4-fluorophenyl)-2, 3-dihydroxypropanoic acid in *SgTAM*, suggesting a similar mechanism for both enzymes (**Figure I.21**).

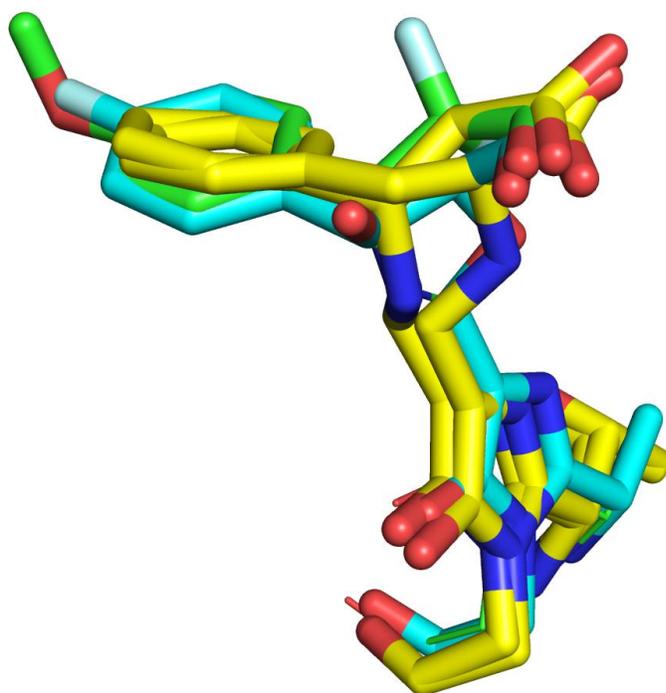


Figure I.21 Overlay of “subunit A” MIO active site of PaPAM with *SgTAM* reveals identical trajectories of the bound ligands. The MIO, (3S)- β -phenylalanine, and (2S)- α -phenylalanine of PaPAM is highlighted in yellow. The MIO of *SgTAM* with bound inhibitor (2S,3S)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid (PDB ID 2RJR) is highlighted in cyan and the MIO of *SgTAM* with bound substrate mimic (3R)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (PDB ID 2RJS) is highlighted in green.

The relationship between the structure of *TcPAM* and *PaPAM* can give clues to the preference between the two enantiomeric forms of β -Phe that are produced by the two enzymes (**Figure I.22**). The carboxylate of the (3S)- β -phenylalanine in *PaPAM* forms a salt bridge to the highly conserved Arg323 (Arg325 in *TcPAM*, Arg311 in *SgTAM* and Arg303 in *RsTAL*) of “subunit C”. Unlike *TcPAM* where both oxygens of the carboxylate are involved in the salt bridge, the interaction is to only one of the oxygens within the carboxylate, allowing the ligand to tilt down towards the MIO (**Figure I.23**). This is significant in *TcPAM*, as it has been theorized that the substrate must rotate in order to present the proper side of the propenoate moiety for amino addition to the opposite side of the substrate. The added interactions of the Arg325 in *TcPAM* likely aids in holding the carboxyl group from rotating while the rest of the substrate rotates about the carboxylate-carbon-carbon bond. In *PaPAM* this rotation is not necessary as the proper face of the beta carbon is positioned towards the MIO after amino excision (**Figure I.24**) in which case closer contact to the MIO is likely preferable. In contrast, the improper positioning of the β -carbon of cinnamate in *TcPAM* prevents addition of the amino group until after phenyl ring rotation.

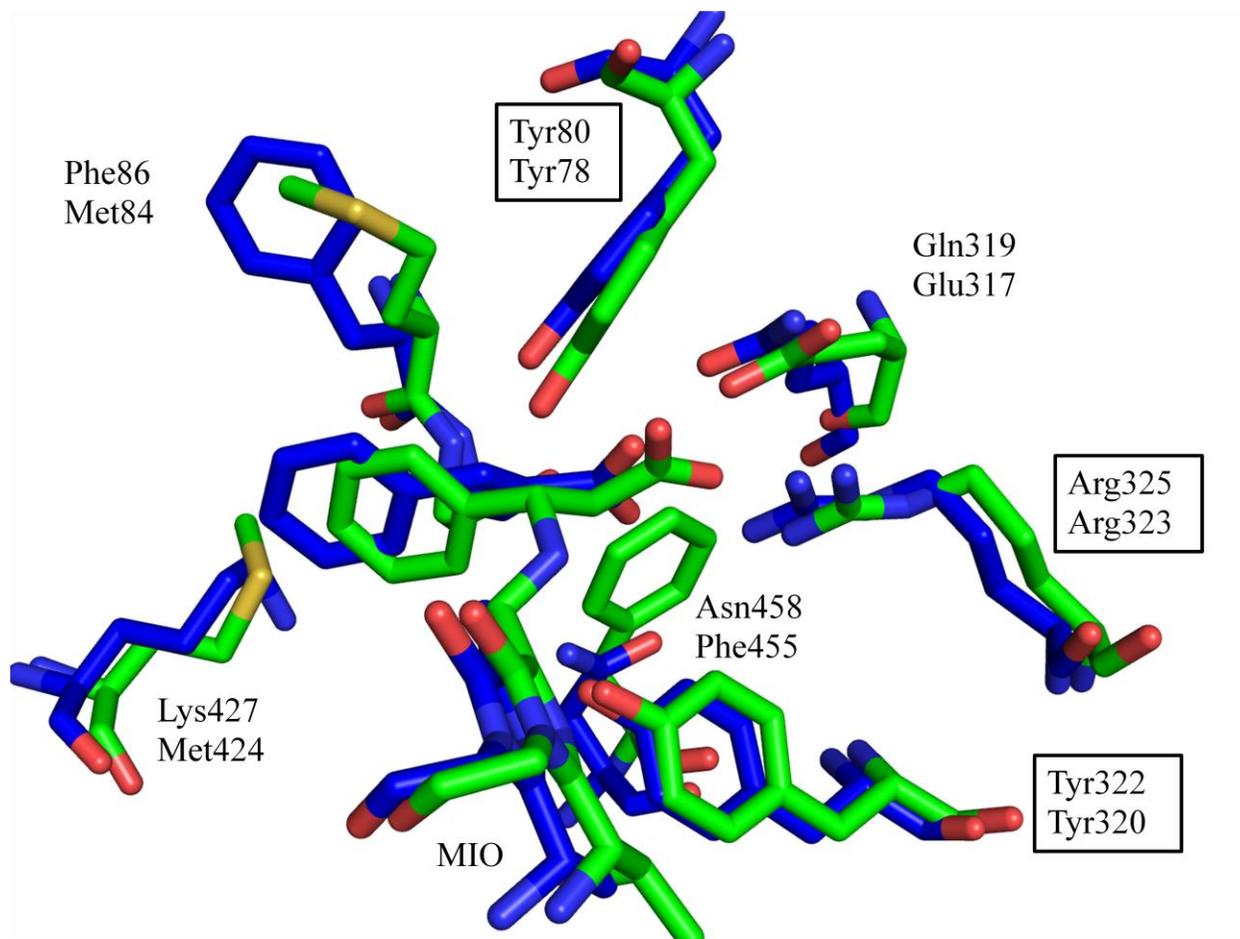


Figure I.22 A comparison of the active site residues of *PaPAM* with (3S)- β -phenylalanine bound (green, PDB ID 3UNV) and *TcPAM* with cinnamate bound (blue, PDB ID 3NZ4) reveals conserved positions for key catalytic residues Tyr80/Tyr78, Tyr322/Tyr320 and Arg325/Arg323. Residues listed on top refer to *PaPAM* and those listed below are that of *TcPAM*.

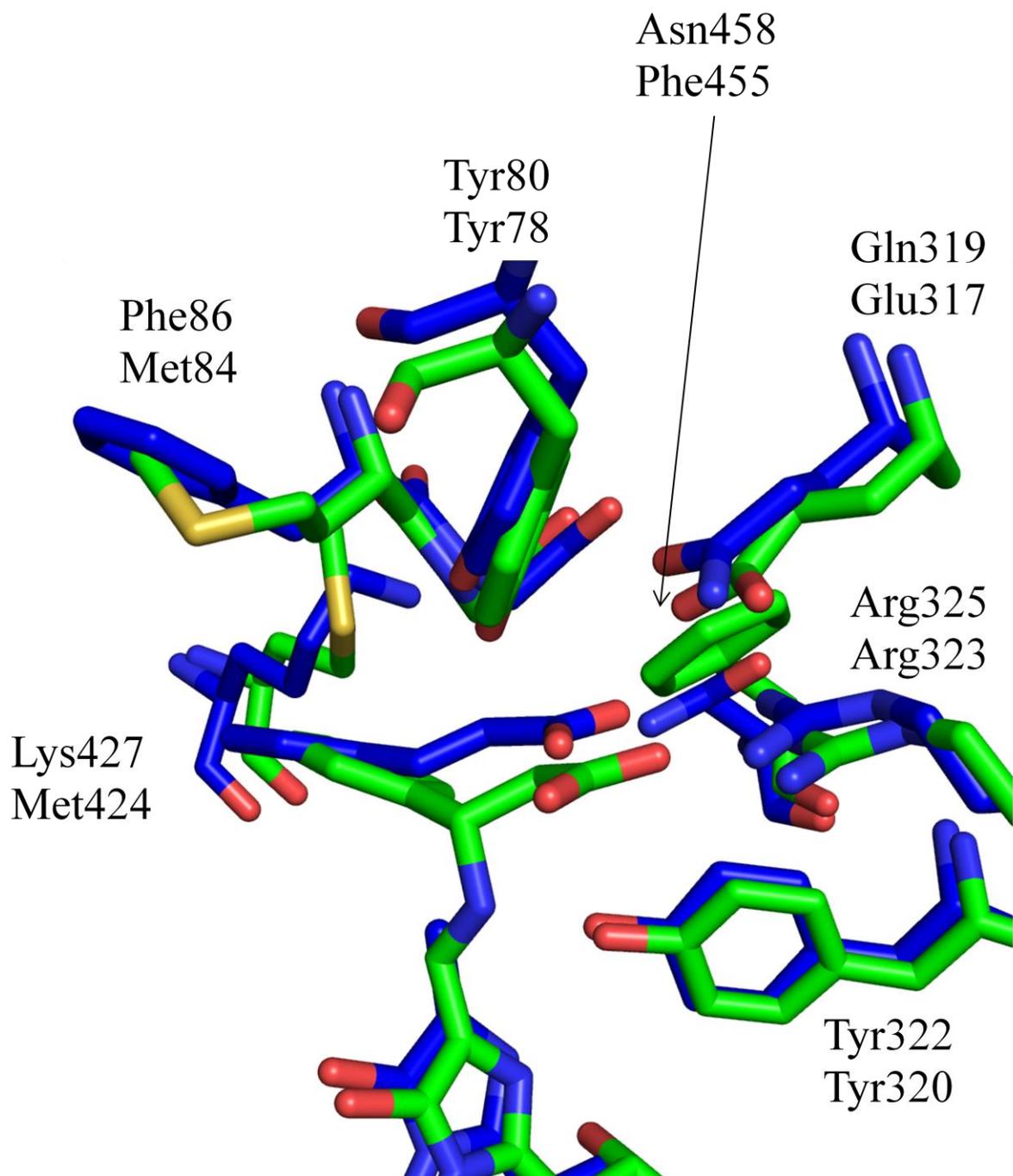


Figure I.23 Comparison of the placement of the ligands within the active sites of PaPAM (green, PDB ID 3UNV) and TcPAM (blue, PDB ID 3NZ4) shows the cinnamate in TcPAM is lifted slightly away from the MIO as compared to the (3S)- β -phenylalanine bound to the MIO of PaPAM. This suggests there is slightly more room above the ligand in TcPAM, perhaps allowing for rotation of the ligand which would allow for stereochemistry observed.

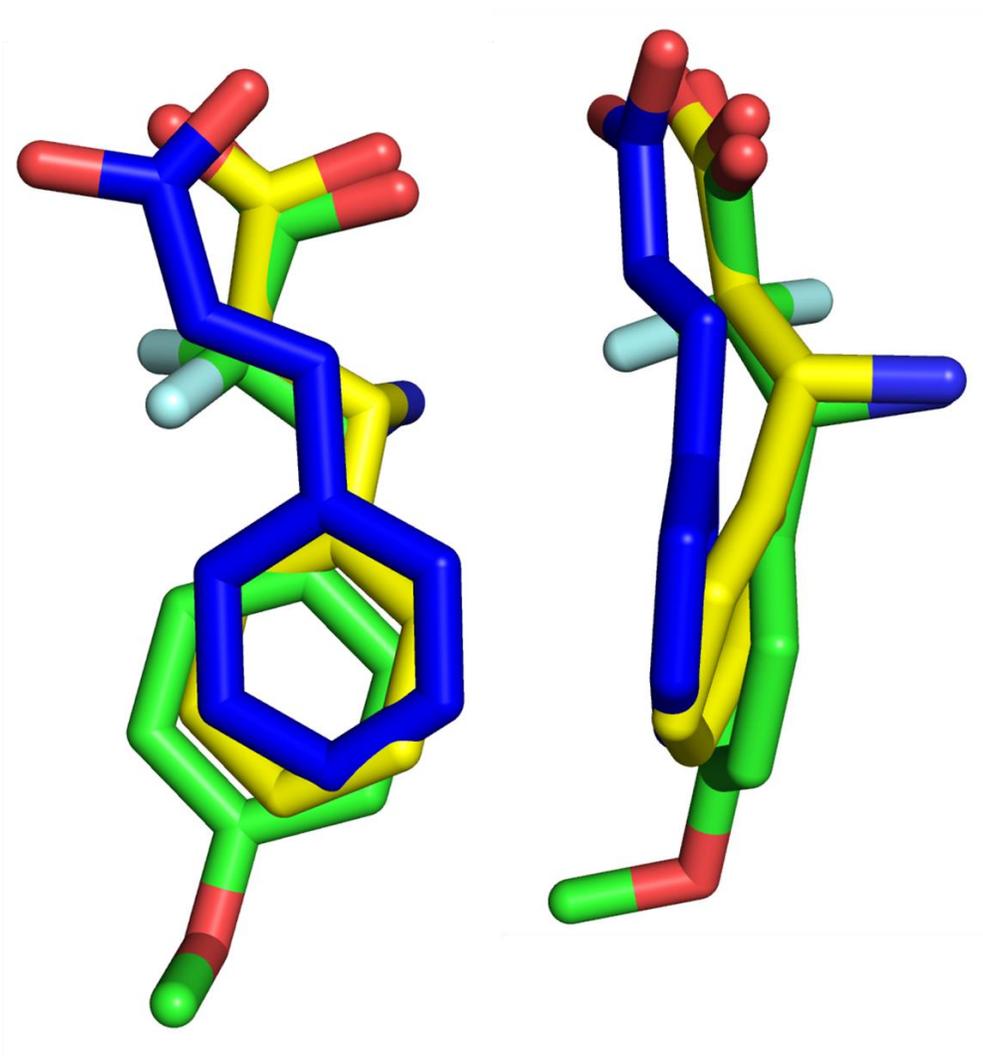


Figure I.24 Overlays of *PaPAM* “subunit C” MIO active site with (3*S*)- β -phenylalanine bound (yellow), *TcPAM* (PDB ID 3NZ4) with cinnamate bound (blue), and *SgTAM* (PDB ID 2RJS) with (3*R*)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (green) bound highlight the variance in the position of the β -carbon within the propenoate moiety which is due largely to the hydrogen bonding interactions with a conserved, flexible Arg that positions the carboxylates.

Also of distinction is Phe455 which differs from the smaller Asn458 found in TcPAM. The Phe455 residue interacts with the phenylalanine, pushing the carboxylate away from Arg323, preventing a double salt bridge interaction and sterically preventing the rotation of the phenyl ring (**Figure I.25**). This conceivably prevents the aryl ring from rotating into the alternative conformation necessary for the production of (3R)- β -phenylalanine (as in TcPAM), thus leading to the production of (3S)- β -phenylalanine exclusively. This architecture which is similar to SgTAM coupled with the presence of the (3S)- β -phenylalanine amino group's direct interaction with the methyldene group of the MIO is highly suggestive of an amino-group alkylation pathway. When the presence of both (2S)- α -phenylalanine and (3S)- β -phenylalanine attached to the MIO is taken into consideration with that stated above, it conclusively proves that bacterial PaPAM does not use a Friedel Crafts mechanism, but rather an amino-group alkylation pathway.

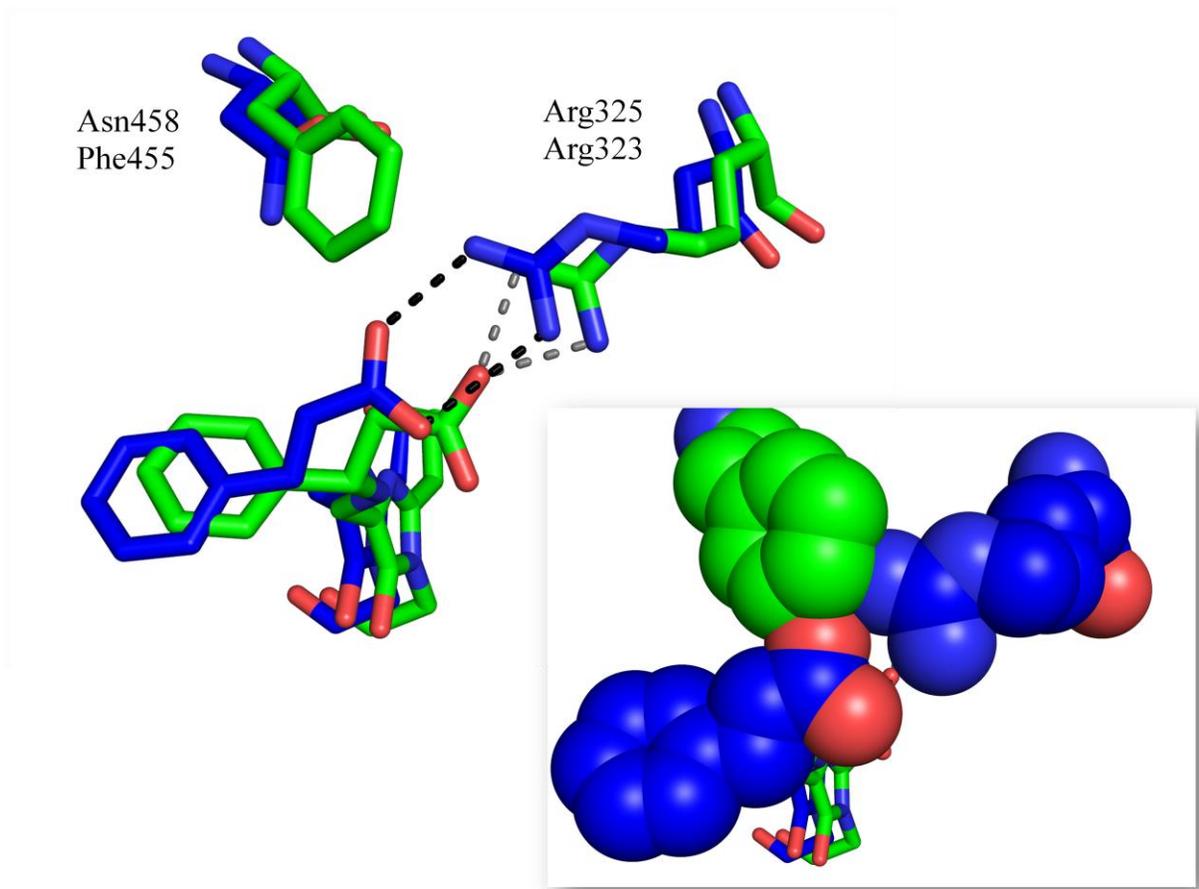


Figure I.25 Top: Interactions of (3S)- β -phenylalanine in *PaPAM* with Arg323 (green, PDB ID 3UNV) are compared to those of *TcPAM* with cinnamate bound to Arg325 (blue, PDB ID 3NZ4). Dark hash marks represent the hydrogen bonds to Arg325 in *TcPAM* as opposed to the grey hash marks representing the hydrogen bonds to Arg323 in *PaPAM* which allows the carboxylates of the ligands to occupy different positions. Bottom: The same as above but with space filling representations for Phe455 in *PaPAM*, Arg325 in *TcPAM* and cinnamate in *TcPAM* showing the collision of the cinnamate ligand to Phe455. Such a collision shows the impact Phe455 has in *PaPAM* in positioning the ligand away from Arg323 such that the nature of the hydrogen bonds is now different. This difference may be important in determining the stereoselectivity of the reaction.

Mutating Phe455 in PaPAM to Asn might therefore switch the stereochemistry of PaPAM to that of TcPAM. However, this point mutation did not affect the stereochemistry of the reaction but did increase the production of trans-cinamate; a result which confirms the importance of Phe455 in properly positioning the ligand for reaction (51). When compared to SgTAM which has the same stereochemistry of PaPAM there is already an Asn (Asn441) in that position (**Figure I.26**). However, additional hydrogen bonding interactions in SgTAM to the hydroxyl group of the phenyl ring might aid in holding the ligand in place. As such interactions do not exist in PaPAM, the steric interactions of Phe455 might be critical. It is important to note that in TcPAM there is also a hydrogen bonding interaction between Asn458 and Glu455. This interaction does not exist in either PaPAM (since the corresponding residues are Phe455 and Thr452) or SgTAM (since the corresponding residues are Asn441 and Asn438). Therefore multiple mutations might be necessary between these two positions to change the stereochemistry of the reaction, particularly a Phe455 to Asn mutation coupled with a Thr452 mutation to Glu to fully mimic the TcPAM active site.

Sequence alignments show EncP from *Streptomyces maritimus*, which was initially characterized as a slow PAL but subsequently shown to have PAM activity with (S)- β -phenylalanine as the product, has an active site similar to PaPAM including a Phe analogous to Phe455 in PaPAM (5, 52-54). As the substrates and active site architectures are similar, this lends more weight to the suggestion the Phe455 is critical in positioning the substrate for PAM activity when (S)- β -phenylalanine is the product. Based on the sequence of PaPAM a BLAST search (55) produced several possible PAMs, particularly a protein from Vibrionales bacterium was identified as having architecture similar to PaPAM and concurrently is known to produce Andrimid (**Table I.3**) (51).

Table I.3 Multiple sequence alignment using Clustal 2.1(32) of proteins identified by a BLAST search (55) as having high sequence similarity to PaPAM suggests four yet to be characterized proteins [*Vibrionales bacterium* SWAT-3 (V. bact), *Bacillus Subtlis* (B.subtl), *Klebsiella pneumoniae* 342 (K.pneu) and *Burkholderia rhizoxinica* (B.rhiz)] as possible PAMS with stereoselectivity analogous to that of PaPAM. *EncP* from *Streptomyces maritimus* is abbreviated S. marit. Potentially important active site residues are highlighted.

PaPAM	MSIVNESGSQPVVSRDETLSQIERTSFHISSGKDISLEEIARAA-RDHQPVTLLH-DEVVN	58
*V.bact	MNIVNEHCKKPVQDSNENLPHADMTSFHLVSGQEVTLDAIAHAA-RHHCPTVD-DGIIQ	58
*S.marit	-----MTFVIELDMNVTLTDQLEDAA-RQRTPVELS-APVRS	34
*B.subtl	-----MENYSFKKFVLS-NQKISLSDFIKIVKEPDLKVEID-DEVKN	40
*K.pneu	-----MNITQHNSTSTGDTFILSPGRNVSLKDFIEFS-QFSKKIVAS-EETRE	46
*B.rhiz	-----MNNMLEITGERIRASDIARVAYDFDIHVRLG-QKACE	36
SgTAM	-----GSMALTQVETEIVPVSVVDGETLTVEAVRVA-EERATVDVP-AESIA	45
AvPAL	--MKTLSQAQSKTSSQQFSFTGNSSANVIIGNQKLTINDVARVARNGTLVSLTNNTDILQ	58
TcPAM	MGFAVESRSRVKDI LGLINTFNVEVKKITVDGTTPI TVAHVAALARRHDVKVALEAEQCRA	60
	:	
PaPAM	RVTRSRSILESMVSDERVIYGVNTSMGGFVNYIVPIAKASELQNNLINAVATNVGKY---	115
*V.bact	RVTASRHILEGMVSDDRVIYGVNTSMGGFVNYIVPIDKASELQNNLIHAVATNVGEY---	115
*S.marit	RVRASRDVLVKFVQDERVIYGVNTSMGGFVDHLVPSQARQLQENLINAVATNVGAY---	91
*B.subtl	KILASRKLLEDEYVENGRIIYGVTTSMGGFVDYLVVPEFSEKLQNNLISSVASNVGEY---	97
*K.pneu	RIAASRRALEKLVKEGSVIYGVNTSMGGFVDHLVPLERAEELQKNLIRGVATNVGER---	103
*B.rhiz	SIVASRKLDDLLQGVKIYGVNTSMGGFVKYLIPEKYATQTQENLIAAVATNVGPY---	93
SgTAM	KAQKSREIFEGIAEQNIPIYGVTTGYGEMIMQVDKSKEVELQTNLVRSHSAGVGPL---	102
AvPAL	GIQASCDYINNAVESGEPIYGVTSFGFGMANVAISREQASELQTNLVWFLKTGAGNK---	115
TcPAM	RVETCSSWVQRKAEDGADIYGVTTGFG--ACSSRRTNQLSELQESLIRCLLAGVFTKGA	118
 * * * * : * . * : . . .	
PaPAM	-----FDDTTVRATMLARIVSLSRGNSAISIVNFKKLIEIYNQGIVPCPEKGSLSGTS	170
*V.bact	-----FDDTTVRATMLARIVSLSRGNSAISIVNFQKLIDIYNRGVVPCVPEKGSLSGTS	170
*S.marit	-----LDDTTARTIMLSRIVSLARGNSAITPANLDKLVAVLNAGIVPCPEKGSLSGTS	146
*B.subtl	-----MSDEDVRATMLARLISLSKGASISLENFKIFLNMLNKNVIPCIPKKGSLGASG	152
*K.pneu	-----FSDIICRATMFARIISLSRGNSALSLENFDRFIAIYNAGLIPPEIPRKGSLGTS	158
*B.rhiz	-----FDDSVVRATMLTRINSLARGVSAISLENIQKFVEIFNKGICPCIPQKGSLSGTS	148
SgTAM	-----FAEDEARAIVAARLNTLAKGHS AVRPIILERLAQYLN EGITPAIPEIGSLGASG	157
AvPAL	-----LPLADVRAAMLLRANSHMRGASGIRLELIKRM EIFLNAGVTPYVYEFGSIGASG	170
TcPAM	SSVDEL PATVTRSAMLLRLNSFTYGC SGIRWEVMEALEKLLNSNVSPKVPLRGSV SASG	178
	: * : * : * * . : . : * . : * : * * . : * * .	
PaPAM	LGPLAAIALVCTGQ---WKARYQGEQMSGAMALEKAGISPMELSFKEGLALINGTSAMVG	227
*V.bact	LGPLAAIALVCTGQ---WKARYH GELMSGSEALKKAGIAPMSLSFKEGLALINGTSAMVG	227
*S.marit	LGPLAAIALVCAGQ---WKARYNGQIMPGRQALSEAGVPEMELSYKDGLALINGTSGMVG	203
*B.subtl	LGPLAFIALVGVGK---WKAKFEGEVLTGEEALIKAKIKMKLGYKEGLALINGTSAMAG	209
*K.pneu	LGPLAAMARMLTGE---GNAWFNGERLAAEDILHQLGLAPLELSYKEGLALINGTSCMVA	215
*B.rhiz	LGPLAAIALALTGK---WKVRYRGEIMSASDALRKTNIEPLRLSYKEGLALINGTSAMTG	205
SgTAM	LAPLSHVASTLIGE---GYVLRDGRPVETAQVLAERGIEPLELRFKEGLALINGTSGMTG	214
AvPAL	LVPLSYITGSLIGLDPSFKVDFNGKEMDAPTALRQLNLSPLTLLPKEGLAMMNGTSVMTG	230
TcPAM	LIPLAYIAGLLIGKPSVIARIGDDVEVPAPPEALSRVGLRPFKQLQAKEGLALVNGTSFATA	238
	* * * : : * . : * . : : * * * * * : * * * . .	

Table I.3 (cont'd)

PaPAM	LGVLLEYDEVKRLFDTYLTVTSLSIEGLHGKTKPFEPVHRMK-PHQGQLEVATTIWETLA	286
*V.bact	LGALLYDEVKRLFDTYLTITALSIEGLHGKTKPFEPVHRMK-PHLGQLEVATTVWETLA	286
*S.marit	LGMTVLQAARRLVDRYLQVSALSVEGLAGMTKPFDPVHGVK-PHRGQRQVASRLWEGLA	262
*B.subtl	TGAMVSDGVKQLLGFYEYISALTFEGLATKLPFDPIVHKRK-LHKGQNYFSTKIYNILK	268
*K.pneu	LAALNVIETRSLLEQYASISAFASETLLARIRPFHPDVHQLK-PHTGQQKIAEMIWNLLQ	274
*B.rhiz	LACLMVSDVEKLIKSYESITALALETLKGKRKVFSPLVHEEK-PHRGQQAASANIYNALA	264
SgTAM	LGSLVVGRALEQAQQAIEVITALLIEAVRGSTSPFLAEGHDIARPHEGQIDTAANMRALMR	274
AvPAL	IAANCVYDTQILTAIAMGVHALDIQALNGTNQSFHPFIHNSK-PHPGQLWAADQMISLLA	289
TcPAM	LASTVMYDANVLLLLLVETLCGMFCEVIFGREEFAHPLIHKVK-PHPGQIESAEELLEWLLR	297
	. . . : . : . * * ** : : :	
PaPAM	DSSSLAVNEHEVEKLI AEEM--DGLVKASNHQIEEDAYSIRCTPQILGPVADTLKNIKQTLT	344
*V.bact	DSSSLAVNEHEVEKMI AEEM--EGTVKASNHQIEEDAYSIRCTPQILGPVADSLKHIQTLT	344
*S.marit	DSHLAVNELDTEQTLGEM--GTVAKAGSLAIEEDAYSIRCTPQILGPVVDVLDRIGATLQ	320
*B.subtl	SSKFVIDEAETEANIQEKK--KNVVEHLDNQIEEDAYSLRCSQPILGPLYETVEFASTIIE	326
*K.pneu	GTRLAVDDIQLSSELGSRL--TNSIKQEDMPIEEDAYSIRCTPQILGPVLETIEFVERIVS	332
*B.rhiz	DSNMISSEDDVSKNLRSQLF-DNVIDSVADQIEEDAYSLRCTPQIIIGPIRDAVDYVKCVVE	323
SgTAM	GSGLTVEHADLRRELQKDKKQKDVQRSEIYLQKAYSLRAIPQVVGAVRDTLYHARHKLK	334
AvPAL	NSQLVRDEL DKGKHDYRDHE-----LIQDRYSLRCLPQYLGPIVDGISQIAKQIE	338
TcPAM	SSPFQDLSREYYSIDKCLK-----PKQDRYALRSSPQWLAPLVQTI RDATTTVE	346
	:: : : : . : * : * . ** : : : : :	
PaPAM	NELNSSNDNPLIDQTTEEVFHNGHFHGQYVSMAMDHLNIALVTMMNLANRRIDRFMDKSN	404
*V.bact	NELNSSNDNPLIDQATEDVFHNGHFHGQYVSMAMDHLNIALVTMMNLANRRVDRFMDKSN	404
*S.marit	DELNSSNDNPIVLPEEA EVFHNGHFHGQYVAMAMDHLNIALVTNLANRRVDRFLDKSN	380
*B.subtl	NEINSSSDNPLILPEENDVFHNGHFHGQYISAMADYLSICLTTLSNLSDRRIDRFMDKSN	386
*K.pneu	NELNSSNDNPLITPENQVVFHNGHFHGQYISAAMDYLTIAIITMCNLSDRRTDRLLTSAN	392
*B.rhiz	NELNSSNDNPLVI PKHGDVYHNGHFHGQYISAMADHLSIALVTLNLSDRRIDRFMDKNN	383
SgTAM	IELNSANDNPLFFEGK-EIFHGANFHGQPIAFAMDFVTIALTQLGVLAERQINRVLNRHL	393
AvPAL	IEINSVTDNPLIDVDNQASYHGGNFLGQYVGMGMDHLRYIIGLLAKHLDVQIALLASPEF	398
TcPAM	TEVNSANDNPIIDHANDRALHGANFQGSVAVGFYMDYVRIAVAGLGKLLFAQFTELMIEYY	406
	* : * * . * * * : . * : * * . : : : : .	
PaPAM	SNGLPFLCAE-NAGLRLGLMGGQFMTASITAESRASCMPMSIQSLS-TTGDFQDIVSFG	462
*V.bact	SNGLPAFLCAE-NAGLRLGLMGGQFMTASITAESRASCMPMSIQSLS-TTGDFQDIVSFG	462
*S.marit	SNGLPAFLCRE-DPGLRLGLMGGQFMTASITAETRSLTIPMSVQSLT-STADDFQDIVSFG	438
*B.subtl	SNGLPAFLTKE-NPGLRLGLMGGQFMSTSLTAENRSLCTPLSIQTLT-STGDFQDIVSFG	444
*K.pneu	SNGLPSFLCAE-NGGLRFGMLGGQFMSSSVTAENRSLATPVSIQTLT-TTGDFQDVVSFG	450
*B.rhiz	SNGLPFLCAN-EQGIRLGLMGGQFMSASLASENRSLCVPVSIHSLP-STADDFQDIVSLG	441
SgTAM	SYGLPEFLVSG-DPGLHSGFAGAQPATALVAENRTIG-PASTQSVS-SNGDNQDVVSMG	450
AvPAL	SNGLPPSLLGNRERKVNMGKGLQICGNSIMPLLTFYGNISADRFPTHAEQFNQININSQG	458
TcPAM	SNGLPGNLSLGPDLSDVYGLKGLDIAMAAYSSELQYLANPVTTTHVHS-AEQHNQDINSLA	465
	* * * * * : : * : * : : . . : : . : * : * .	
PaPAM	LVAARRVREQLKLNKYVVSFELLCACQAVDIRG-----	495
*V.bact	LVAARRVREQLKLNKYVVSFELLCACQAADIRG-----	495
*S.marit	FVAARRAREVLTNAAYVVAFELLCACQAVDIRG-----	471
*B.subtl	LIASRRCKEILENTLYIVSFELLCACQAIDIRE-----	477
*K.pneu	LVAARRTAEVLQNTRYVIAFELICAAQAADIRD-----	483
*B.rhiz	LVAARRAQEIFNNTVYVVSFELLCACQAAADIRG-----	474
SgTAM	LISARNARRVLSNKKILAVEYLAQAQAVDISGR-----	484
AvPAL	YTSATLARRSVDIFQNYVAIALMFGVQAVDLRQYKKTGHYDAR-----	501
TcPAM	LISARKTDEALDILKLMIAHSLTAMCQAVDLRQLEALVKVVENNVSTLADECGLPNDTK	525
	:: . . . : * * * :	

Table I.3 (cont'd)

PaPAM	-----T	496
*V.bact	-----T	496
*S.marit	-----A	472
*B.subtl	-----E	478
*K.pneu	-----A	484
*B.rhiz	-----A	475
SgTAM	-----F	485
AvPAL	-----	
TcPAM	ARLLYVAKAVPVYTYLESPCDPTLPLLLGLKQSCFDSILALHKKDGIETDTLVDRLAEFE	585
PaPAM	AGLSKRTRALYDKTRTLVPYLE-----E	519
*V.bact.pro	TGLSTQTRALYERTRTVVPYLE-----Q	519
*S.marit	DKLSSFTRPLYERTRKIVPFFD-----R	495
*B.subtl	SNLSNATKVLVDNVRKIVPYS-----Y	501
*K.pneu	TKLGNSGRLWYAKVRESVPYLD-----H	507
*B.rhiz	DKLGHTAMLYNSVRSFLPFFE-----K	498
SgTAM	DGLSPAAKATYEAVRRLVPTLG-----V	508
AvPAL	ACLSPATERLYSAVRHVVGQK-----PTSDRPYIWNDN	534
TcPAM	KRLSDRLENEMTAVRVLYEKKGHKTADNNDALVRIQGSKFLPFYRFVRDELDTGVMSARR	645
	. .	
PaPAM	DKTISDYIESIAQTVLTKN---SDI-----	541
*V.bact	DHTITDYIEGIAQTVLTKN---HAL-----	541
*S.marit	DETITDYVEKLAADLIAGEPVDAVAHAH-----	523
*B.subtl	DTSITPFIEELKYLKQKTTLLKELDNITSIDINK-----	535
*K.pneu	DESITPYLEELVSRILGGHS-----	527
*B.rhiz	DESLTPYLENIAMFIRNEMAQSLGGD-----	524
SgTAM	DRYMADDIELVADALSRGEFLRAIARETDIQLR-----	541
AvPAL	EQGLDEHIARISADIAAGGVIVQAVQDILPCLH-----	567
TcPAM	EQTPQEDVQKVFDAIADGRITVPLLHCLQGFLGQPNGCANGV	687
	: : : :	

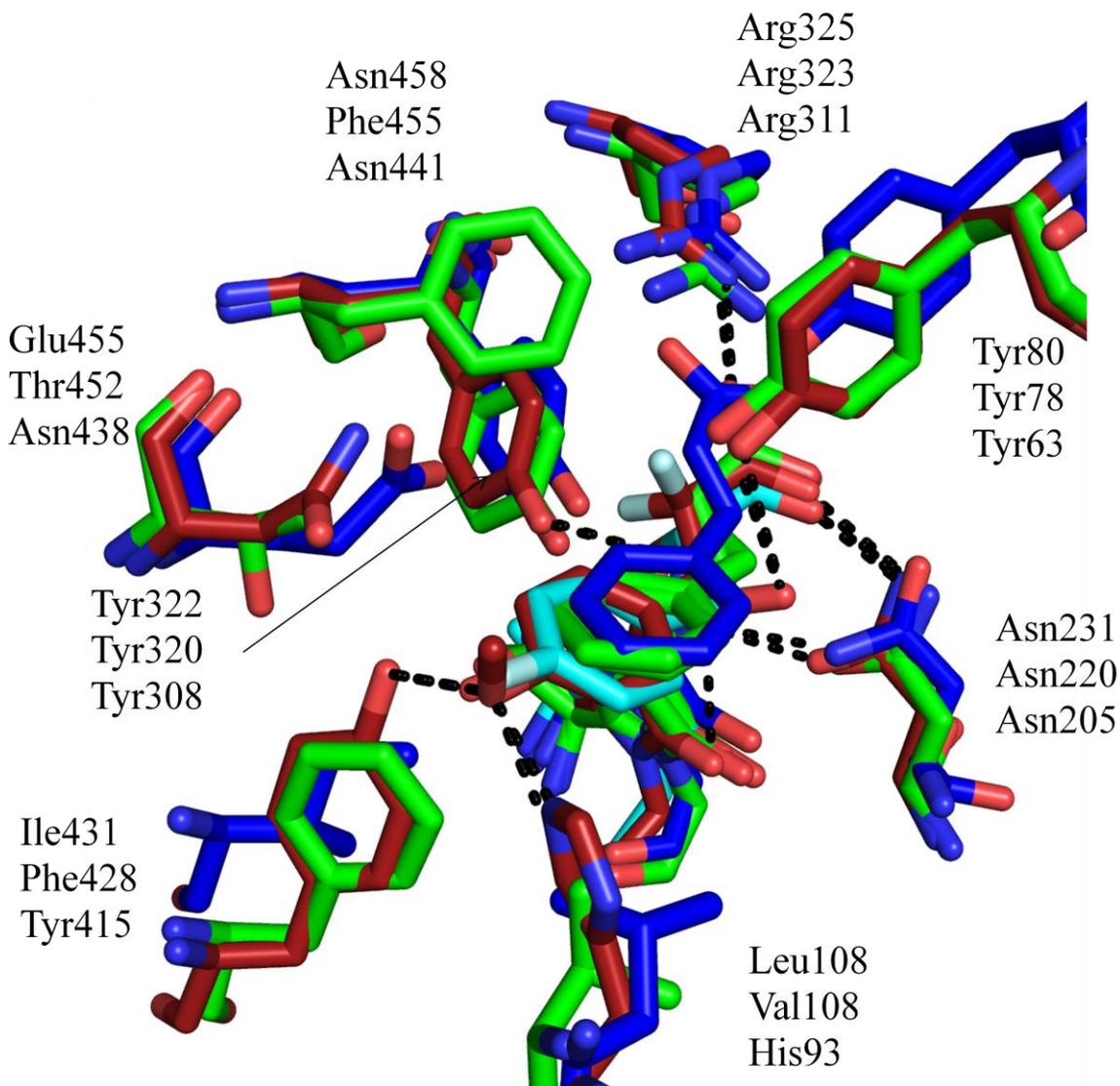


Figure I.26 The active site residues of *PaPAM* with (3*S*)- β -phenylalanine (green), *TcPAM* with cinnamate (blue), and *SgTAM* with (3*R*)-3-amino-2,2-difluoro-3-(4-methoxyphenyl)propanoic acid (red) are overlaid with the ligand (2*S*,3*S*)-3-(4-fluorophenyl)-2,3-dihydroxypropanoic acid (cyan) from *SgTAM* showing that at the location of Phe455 in *PaPAM* are non-conserved Asn residues in *TcPAM* and *SgTAM*. Adjacent to this position is a non-conserved Thr452 which pairs to Glu455 in *TcPAM* and Asn438 in *SgTAM*. These two sites may be responsible together for the stereospecificity of these aminomutases. Residues are listed in the order *TcPAM* (top), *PaPAM* (middle), and *SgTAM* (bottom). Hydrogen bonding interactions for *SgTAM* are shown as black hash lines highlighting the additional connections to the phenyl ring of the ligand. Hydrogen bonding interactions for *TcPAM* are shown in orange hash marks to show the interactions between Glu455 and Asn458 which may be important in stereoselectivity.

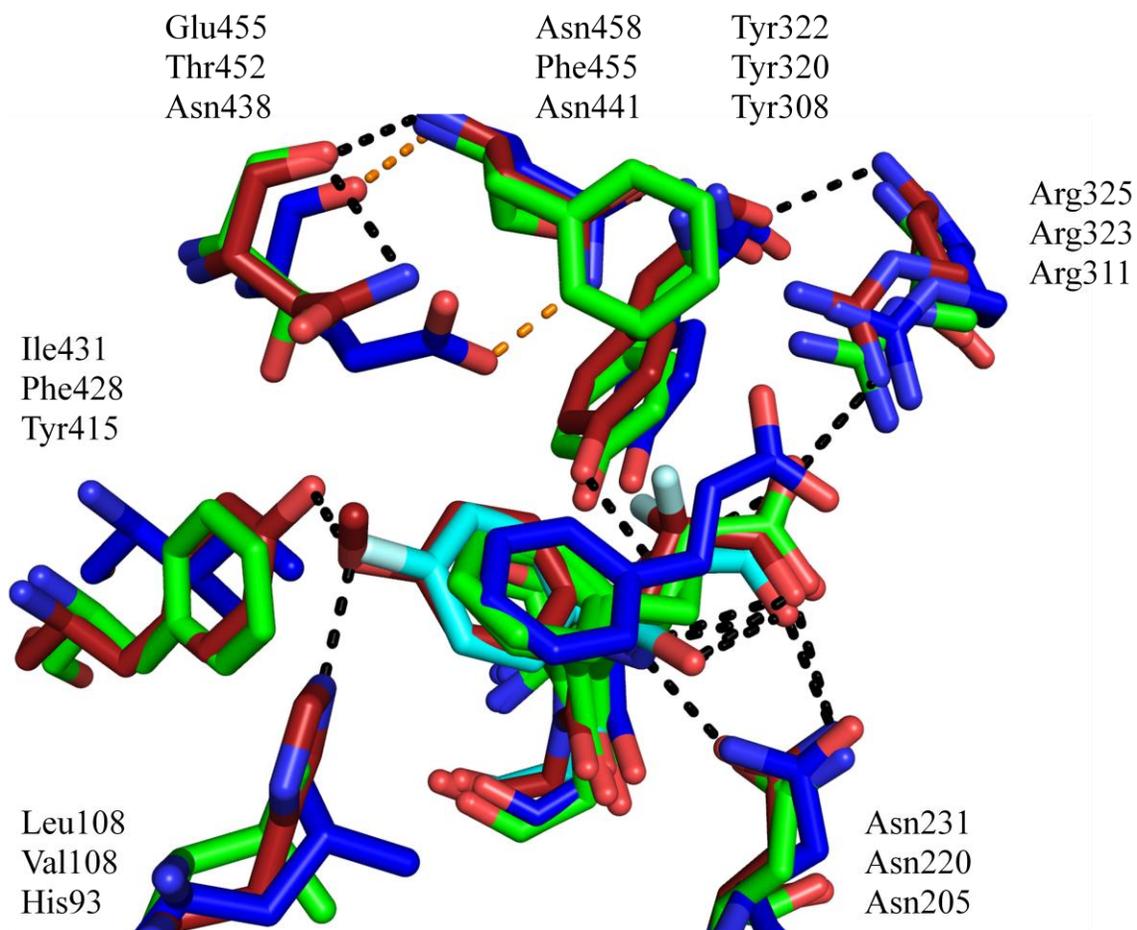


Figure I.26 (cont'd)

REFERENCES

REFERENCES

- 1 Vick J, Schmidt-Dannert C (2011). "Expanding the enzyme toolbox for biocatalysis". *Angew. Chem. Int. Ed.* **50**: 2-5.
- 2 Nestl B, Nebel B, Hauer B (2011). "Recent progress in industrial biocatalysis". *Current Opinion in Chemical Biology.* **15**: 187-193.
- 3 Wu B, Szmanski W, Wijma HJ, Crismaru CG, de Wildeman S, Poelarends GJ, Feringa BL, Janssen DB (2010). "Engineering of an enantioselective tyrosine aminomutase by mutation of a single active site residue in phenylalanine aminomutase". *Chem. Commun.* **46**: 8157-8159.
- 4 Ratnayake N, Wanninayake U, Geiger J, Walker K (2011). "Stereochemistry and Mechanism of a Microbial Phenylalanine Aminomutase". *J. Amer. Chem. Soc.* **133**: 8531-8533.
- 5 Moffitt MC, Louie GV, Bowman ME, Pence J, Noel JP, Moore BS (2007). "Discovery of two cyanobacterial phenylalanine ammonia lyases: Kinetic and structural characterization". *Biochemistry* **46**: 1004–1012.
- 6 Wang L, Gamez A, Archer H, Abola EE, Sarkissian CN, Fitzpatrick P, Wendt D, Zhang YH, Vellard M, Bliesath J, Bell SM, Lemontt JF, Scriver CR, Stevens RC (2008). "Structural and biochemical characterization of the therapeutic *Anabaena variabilis* phenylalanine ammonia lyase". *Journal of Molecular Biology* **380** (4): 623-635.
- 7 Wu B, Szymanski W, Heberling M, Feringa BL, Janssen DB (2011). "Aminomutases: mechanistic diversity, biotechnological applications and future perspectives". *Trends in Biotechnology.* **29** (7): 352-359.
- 8 Walker K, Klettke K, Akiyama T, Croteau R (2004). "Cloning, heterologous expression, and characterization of a phenylalanine aminomutase involved in Taxol biosynthesis". *J. Biol. Chem.* **279**: 53947-53954.
- 9 Jordan MA, Wendell K, Gardiner S, Derry WB, Copp H, Wilson L (1996). "Mitotic block induced in HeLa cells by low concentrations of paclitaxel (Taxol) results in abnormal mitotic exit and apoptotic cell death". *Cancer Res.* **56**: 816-825.
- 10 Nogales E, Wolf SG, Downing KH (1998). "Structure of the $\alpha\beta$ tubulin dimer by electron crystallography". *Nature* **391**: 199- 203
- 11 Gascoigne KE, Taylor SS (2009). "How do anti-mitotic drugs kill cancer cells?". *Jour. Of Cell Science* **122** (15): 2579-2585.

- 12 Holton RA, Somoza C, Kim HB, Liang F, Biediger RJ, Boatman PD, Shindo M, Smith CC, Kim S (1994). "First total synthesis of taxol. 1. Functionaliation of the B ring". *J. Am. Chem. Soc.* **116** (4): 1597-1598.
- 13 Holton RA, Kim HB, Somoza C, Liang F, Biediger RJ, Boatman PD, Shindo M, Smith CC, Kim S (1994). "First total synthesis of taxol. 2. Completion of the C and D rings". *J. Am. Chem. Soc.* **116** (4): 1599-1600.
- 14 Nicolaou KC, Yang Z, Liu JJ, Ueno H, Nantermet PG, Guy RK, Claiborne CF, Renaud J, Couladouros EA, Paulvannan K, Sorensen EJ (1994). "Total synthesis of taxol". *Nature* **367**: 630-634.
- 15 Mountford PG (2010), "The Taxol ® Story – Development of a Green Synthesis via Plant Cell Fermentation". *Green Chemistry in the Pharmaceutical Industry.* **7**:145-160
- 16 Malik S, Cusido RM, Mirjalili MH, Moyano E, Palazon J, Bonfill M (2011). Production of the anticancer drug taxol in *Taxus baccata* suspension cultures: A review". *Process Biochemistry* **46**: 23-34.
- 17 Heinig U, Jennewein S (2009), "Taxol: A complex diterpenoid natural product with an evolutionarily obscure origin". *African Journal of Biotechnology* **8** (8): 1370-1385.
- 18 Guo BH, Kai GY, Jin HB, Tang KX (2005) "Taxol synthesis". *African Journal of Biotechnology.* **5** (1): 015-020.
- 19 U.S. Enviornmental Protection Agency (2010). "2004 Greener Synthetic Pathways Award" Retrieved June 1, 2011, from <http://www.epa.gov/greenchemistry/pubs/pgcc/winners/gspa04.html>
- 20 Klettke KL, Sanyal S, Mutatu W, Walker KD (2007) "β-Styryl- and β-arylb-alanine products of phenylalanine aminomutase catalysis". *J. Am. Chem. Soc.* **129**: 6988-6989.
- 21 Calabrese JC, Jordan DB, Boodhoo A, Sariaslani S, Vannelli T (2004). "Crystal structure of phenylalanine ammonia lyase: Multiple helix dipoles implicated in catalysis". *Biochemistry* **43**: 11403–11416.
- 22 Schwede TF, Retey J, Schulz GE (1999). "Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile". *Biochemistry* **38**: 5355–5361.
- 23 Rother D, Poppe L, Viergutz S, Langer B, Retey J (2001). "Characterization of the active site of histidine ammonia-lyase from *Pseudomonas putida*". *Eur. J. Biochem.* **268**: 6011–6019.
- 24 Wickner RB (1969) "Dehydroalanine in histidine ammonia lyase". *J. Biol. Chem.* **244**: 6550-6552.

- 25 Cooke H, Bruner S (2010). “Probing the active site of MIO-dependent aminomutases, key catalysts in the biosynthesis of β -amino acids incorporated in secondary metabolites”. *Biopolymers*. **93** (9): 802-810.
- 26 Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ (1996) “Crystal structure of the *Aequorea victoria* green fluorescent protein”. *Science*. **273**: 1392-1395.
- 27 Baedeker M, Schulz GE (2002) “Autocatalytic peptide cyclization during chain folding of Histidine Ammonia-Lyase”. *Structure*. **10**: 61-67.
- 28 Hanson K, Havir E (1970). “L-phenylalanine ammonia-lyase. IV. Evidence that the prosthetic group contains a dehydroalanyl residue and mechanism of action”. *Biochem. Biophys.* **141**: 1-17.
- 29 Cooke HA, Christianson CV, Bruner SD (2009) “Structure and chemistry of 4-methylideneimidazole-5-one containing enzymes”. *Current Opinion in Chemical Biology*. **13**: 460-468.
- 30 Turner N (2011). “Ammonia lyases and aminomutases as biocatalysts for the synthesis of α -amino and β -amino acids”. *Current Opinion in Chemical Biology*. **15**: 234-240.
- 31 Schuster B, Retey J, (1995) “The mechanism of action of phenylalanine ammonia-lyase: the role of prosthetic dehydroalanine” *Proc. Natl. Acad. Sci. USA* **92**: 8433 – 8437.
- 32 Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998). “Multiple sequence alignment with Clustal X”. *Trends Biochem. Sci.* **23**: 403-405.
- 33 Maïke P, Joachim H, Ulrich M (2010) “Biosynthesis of phenylpropanoids and related compounds”. *Annual Plant Reviews: Biochemistry of Plant Secondary Metabolism* **40**: 182-257.
- 34 Vogt T (2010) “Phenylpropanoid Biosynthesis”. *Molecular Plant* **3**(1): 2-20
- 35 Wang L, Gamez A, Archer H, Abola EE, Sarkissian CN, Fitzpatrick P, Wendt D, Zhang Y, Vellard M, Bliessath J, Bell SM, Lemontt JF, Scriver CR, Stevens RC (2008) “Structural and biochemical characterization of the therapeutic *Anabaena variabilis* phenylalanine ammonia lyase. *J. Mol. Biol.* **380**: 623-635.
- 36 Rother D, Poppe L, Morlock G, Viergutz S, Retey J (2002). “An active site homology model of phenylalanine ammonia-lyase from *Petroselinum crispum*”. *Eur. J. Biochem.* **269**: 3065–3075.
- 37 Ritter H, Schulz GE (2004). “Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase”. *Plant Cell* **16**: 3426–3436.

- 38 Bartsch S, Bornscheuer UT (2009). "A single residue influences the reaction mechanism of ammonia lyases and mutases". *Angew. Chem. Int. Ed.* **48**: 3362–3365.
- 39 Louie GV, Bowman ME, Moffitt MC, Baiga TJ, Moore BS, Noel JP (2006). "Structural determinants and modulation of substrate specificity in phenylalanine-tyrosine ammonia-lyases". *Chem. Biol.* **13**: 1327–1338.
- 40 Christenson SD, Liu W, Toney MD, Shen B (2003) "A novel 4-methylideneimidazole-5-one-containing tyrosine aminomutase in enediyne antitumor antibiotic C-1027 biosynthesis". *J. Am. Chem. Soc.* **125**: 6062-6063.
- 41 Christianson CV, Montavon TJ, Van Lanen SG, Shen B, Bruner SD (2007). "The structure of L-tyrosine 2,3-aminomutase from the C-1027 enediyne antitumor antibiotic biosynthetic pathway". *Biochemistry* **46**: 7205–7214.
- 42 Montavon TJ, Christianson CV, Festin GM, Shen B, Bruner SD (2008). "Design and characterization of mechanism-based inhibitors for the tyrosine aminomutase SgTAM". *Bioorg. Med. Chem. Lett.* **18**: 3099–3102.
- 43 Feng L, Wanninayake U, Strom S, Geiger J, Walker K (2011). "Mechanistic, mutational, and structural evolution of a *Taxus* phenylalanine aminomutase". *Biochemistry.* **50**: 2919-2930.
- 44 Mutatu W, Klettke KL, Foster C, Walker KD (2007) "Unusual mechanism for an aminomutase rearrangement: retention of configuration at the migration termini". *Biochemistry.* **46**: 9785-9794.
- 45 Magarvey NA, Fortin PD, Thomas PM, Kelleher NL, Walsh CT (2008) "Gatekeeping versus Promiscuity in the Early Stages of the Andrimid Biosynthetic Assembly Line". *ACS Chem. Biol.* **19**: 3(9): 542-554.
- 46 Arnold K, Bordoli L, Kopp J, Schwede T (2006). "The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling". *Bioinformatics* **22**:195-201.
- 47 Collaborative Computational Project, Number 4 (1994). "The CCP4 Suite: Programs for Protein Crystallography". *Acta Cryst.* **D50**: 760-763.
- 48 Lebedev A (2011) "Jligand" Retrieved Fri Jun 3, 2011, from <http://www.ysbl.york.ac.uk/mxstat/JLigand/index.html>
- 49 Liu X, Fortin PD, Walsh CT (2008). "Andrimid producers encode an acetyl-CoA carboxyltransferase subunit resistant to the action of the antibiotic". *PNAS* **105** (36): 13321-13326.

- 50 Jin M, M. Fischbach MA, Clardy J (2006). "A Biosynthetic Gene Cluster for the Acetyl-CoA Carboxylase Inhibitor Andrimid". *JACS* **128**: 10660-10661.
- 51 Strom S, Wanninayake U, Ratnayake ND, Walker KD, Geiger JH (2012). "Insights into the Mechanistic Pathway of the *Pantoea agglomerans* Phenylalanine Aminomutase". *Angew. Chem. Int. Ed.* **51**: To be Published
- 52 Xiang LK, Moore BS (2002). "Inactivation, Complementation and Heterologous Expression of encP, a Novel Bacterial Phenylalanine Ammonia-Lyase Encoding Gene". *J. Biol. Chem.* **277**: 32505-32509.
- 53 Xiang LK, Moore BS (2005). "Biochemical Characterization of a Prokaryotic Phenylalanine Ammonia-Lyase". *J. Bacteriol.* **187**: 4286-4289. Correction: (2006) *J. Bacteriol.* **188**: 5331.
- 54 Grecu T, Masters of Philosophy Thesis, University of Manchester (UK), 2010.
- 55 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). "Basic local alignment search tool". *J. Mol. Biol.* **215**:403-410.
- 56 Otwinowski Z, Minor W (1997). "Processing of X-ray Diffraction Data Collected in Oscillation Mode". *Methods in Enzymology.* **276** (Macromolecular Crystallography, part A): 307-326. C.W. Carter, Jr. & R. M. Sweet, Eds., Academic Press (New York).

Chapter II: BadA

II.1. Background

II.1.1. Coenzyme A Ligases Structure and Function

Benzoate-Coenzyme A (CoA) Ligases (BCLs) are acyltransferases that facilitate the addition of acetyl CoA to benzoate to produce benzoyl CoA (1). They are members of the PFAM0501 adenylate-forming family of ligases (2, 3) which includes acyl-coenzyme A ligases, firefly luciferases and peptide synthetases. Members of this family are capable of carrying out a two step reaction. First, they adenylate carboxylate containing compounds such as benzoates and fatty acids. The adenylate is then reacted with acetyl CoA to produce acetyl-CoA derivatives. The production of these acetyl-coenzyme A acids activates the aromatic ring or long alkyl chain for biological degradation in certain anaerobic bacteria (4). Germaine to our investigation, in the Taxol biosynthetic pathway benzoyl CoA is used in the final synthetic step by N-debenzoyl-2'-deoxytaxol N-benzoyltransferase (DBTNBT) to benzamidate the side chain of Taxol (**Figure II.1**) (5).

Additionally, coenzyme A ligases can be used biosynthetically to produce a host of new acyl-CoA derivatives. One ligase known for its promiscuity is the *Rhodospseudomonas palustris* Benzoate-Coenzyme A Ligase (BadA) (6). *Rhodospseudomonas palustris* is a non-sulfur purple phototrophic bacterium capable of using both light and aromatic compounds as energy sources (7). It was studied by Harwood and Gibson for its anaerobic degradation of [^{14}C]benzoate, during which they identified BadA as the coenzyme A ligase responsible for the efficient break down of benzoate as a carbon source; a typical first step in the benzoate degradation pathway (for a review see 8). Though benzoic acid is its primary substrate, BadA can additionally accept certain ortho substituted benzoates. As with other members of this family it first adenylates them with ATP, reacting the acyl-adenylate with CoA to form benzyl-CoA via thioesterification. Encouraged by its current promiscuity, knowing the structure of the active site could allow for the production of mutants capable of producing meta and para substituted benzoates as well. When incorporated into known biosynthetic pathways for Taxol such mutants could be used to produce Taxol derivatives (9).

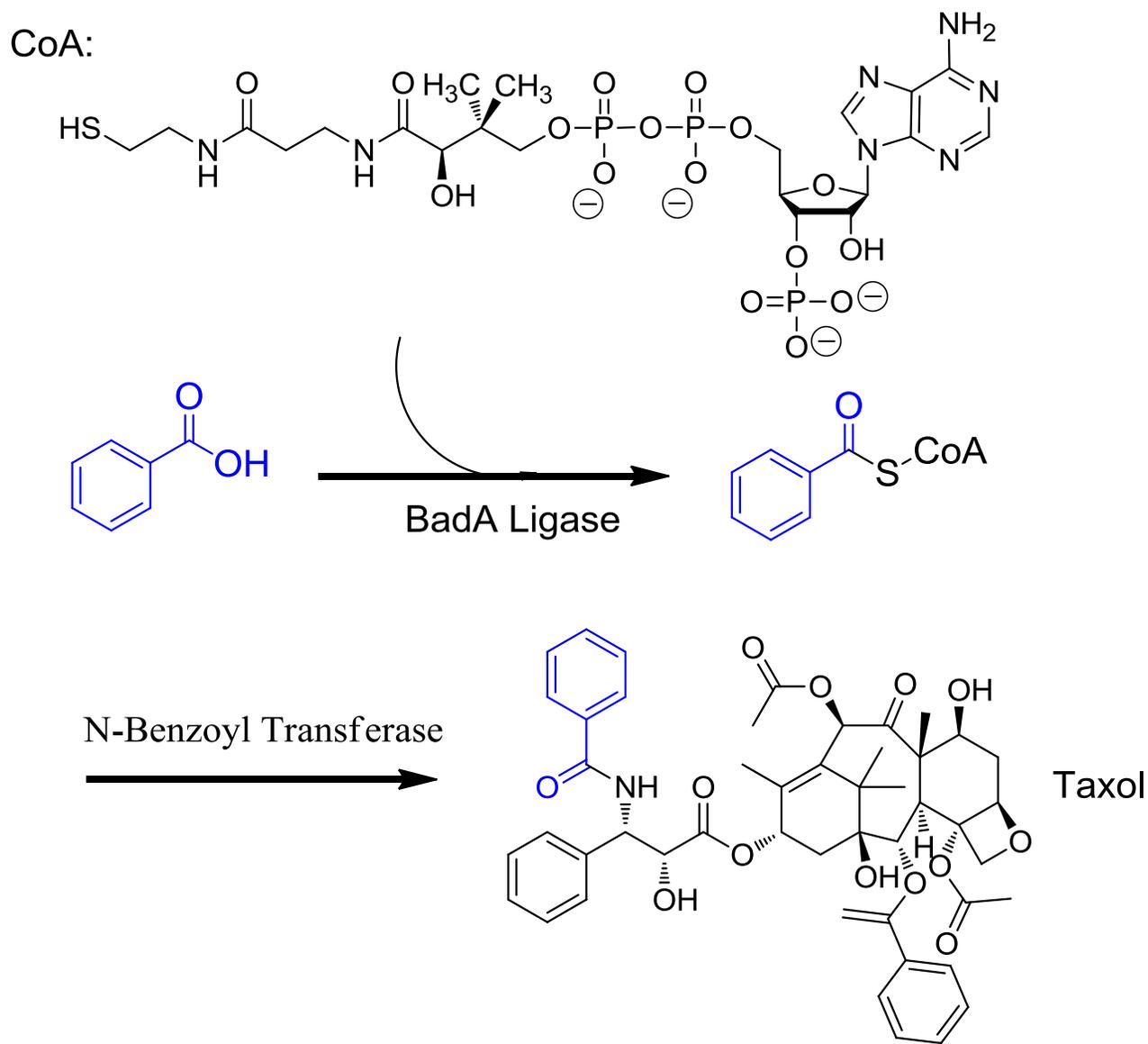


Figure II.1 The ligation of Coenzyme A to benzoic acid by the *Rhodospseudomonas palustris* Benzoate-Coenzyme A ligase (BadA) is an important step in producing benzyl-CoA, a substrate used biosynthetically to produce small molecules such as Taxol.

II.1.2. Benzoate and Benzoate derivatives Coenzyme A Ligase Structures

The overall architecture of the Coenzyme A ligases consists of a concise N-terminal domain and a smaller C-terminal domain with the two being connected by a short, flexible hinge region (**Figure II.2**). The active site of the enzyme rests at the interface of these two domains near the hinge region.

There are two different conformations of the enzyme believed to be individually important for the adenylation and thioesterification reactions. This movement has been extensively studied in the Human Medium-chain Acyl-coenzyme A Synthetase ACSM2A (ACSM2A) L64P mutant (10). The structure of this synthetase has been solved both apo (PDB ID 3B7W) and with a variety of ligands bound to the active site including the products Butyryl Coenzyme A and AMP (PDB ID 3EQ6), ATP (PDB ID 3C5E), AMP (PDB ID 2VZE), CoA (PDB ID 3GPC), AMP-CPP non-hydrolysable ATP (PDB ID 3DAY), and ibuprofen (PDB ID 2WD9). These structures demonstrate that the two different conformations are affected by the ligand bound to the enzyme.

The first conformation is referred to as the “adenylation conformation” as it is required for the adenylation of the substrate carboxylate by adenosine triphosphate (ATP). In the case of ACSM2A this step requires magnesium ion which is observed in the crystal structure to interact with the phosphates of ATP (PDB ID 3C5E). The second conformation, the “thioesterification conformation”, is required for reaction of the adenylated substrate with CoA. The two are related by a large, essentially rigid movement of the C-terminal domain in relation to the N-terminal domain.

The movement has also been studied in the x-ray crystal structure of 4-Chlorobenzoate: Coenzyme A Ligase (CBL), the structure of which has been determined with several ligands including 4-chlorobenzoate, 4-chlorobenzoyl-AMP and 4-chlorophenacyl CoA (PDB ID 1T5D, 3CW8 and 3CW9 respectively) (**Figure II.2**) (11). The percent identity between BadA and CBL is 27%, with the reactions being almost identical. The low percent identity is common among members of this class of enzymes. These structures of CBL highlight the residues critical for enzymatic activity (**Figure II.3**).

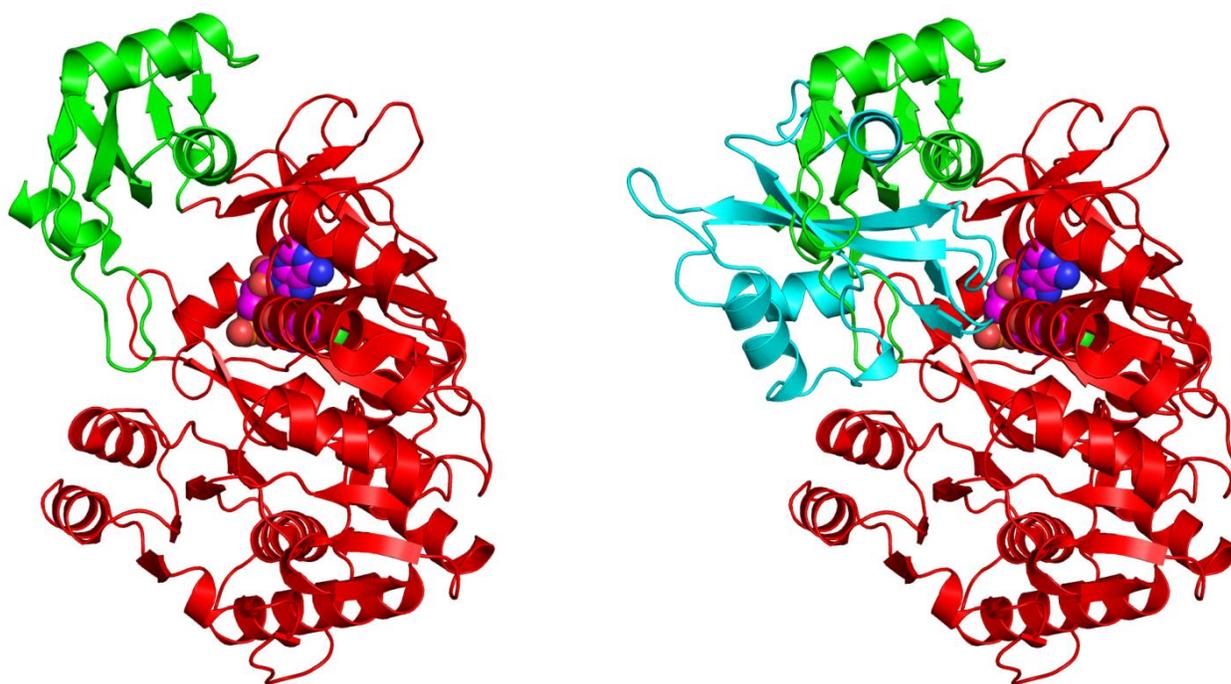


Figure II.2 The adenylation and thioesterification conformations are related by a large C-terminal domain movement in this class of benzoate CoA ligases. Left: CBL (N-terminal domain: red; C-terminal domain: green) with 4-chlorobenzoyl-AMP (magenta spheres) bound in the active site at the interface of the two domains. Right: Overlay of CBL with BCL_M (cyan) demonstrates the perfect overlap of the N-terminal domains compared to the large domain movement of the C-terminal domains.

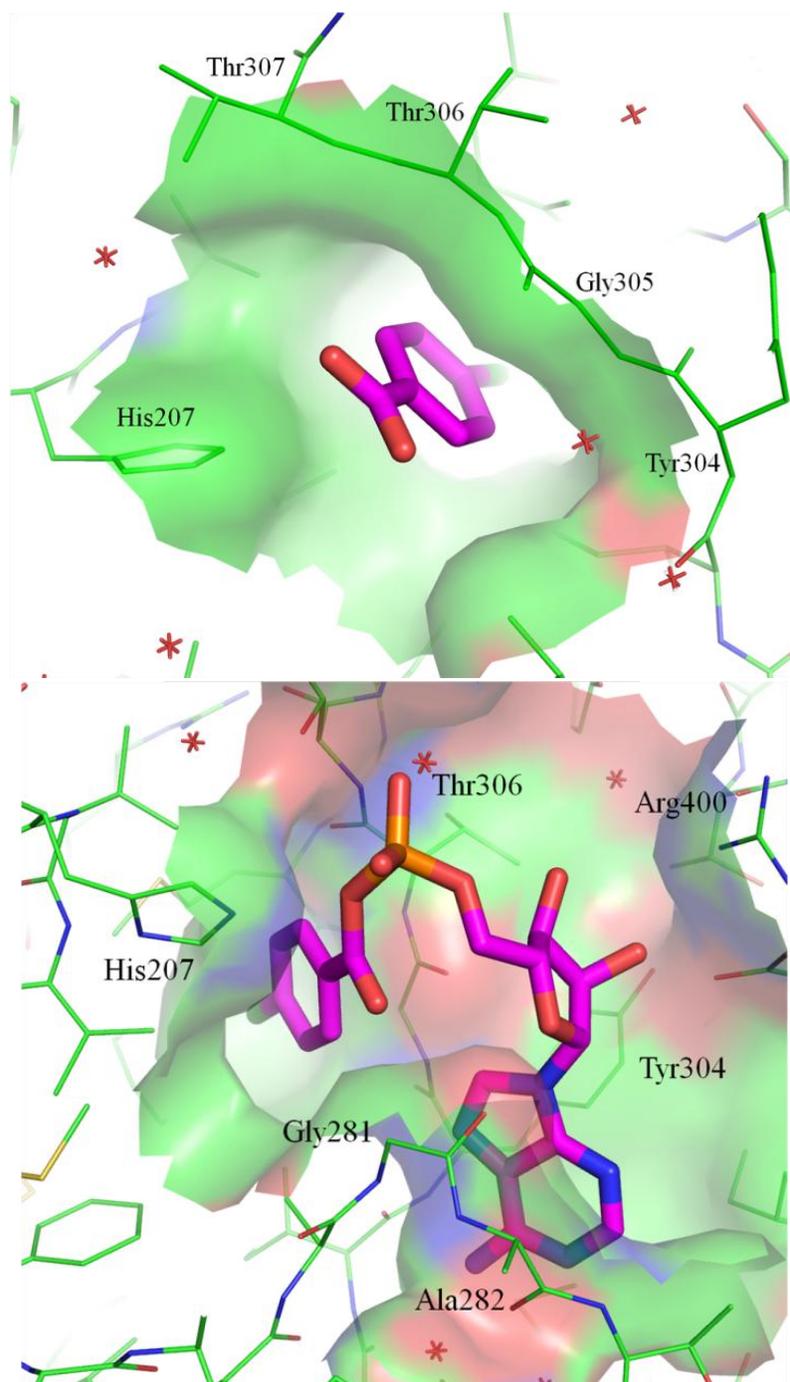


Figure II.3 The various arrangements of the active site are highly affected by the ligand attached in CBL. Top: 4-chlorobenzoate bound showing the openness of the binding pocket towards the para position of the phenyl ring. Bottom: 4-chlorobenzoyl-AMP bound. The ATP binding pocket is just adjacent to the 4-chlorobinding pocket. Next page: 4-chlorophenacyl CoA bound showing the CoA channel which extends to the protein surface (left side of image). Ligands are highlighted in magenta. The protein residues are shown in green. Waters appear in red. The colored mesh represents the surface area the ligand is exposed to.

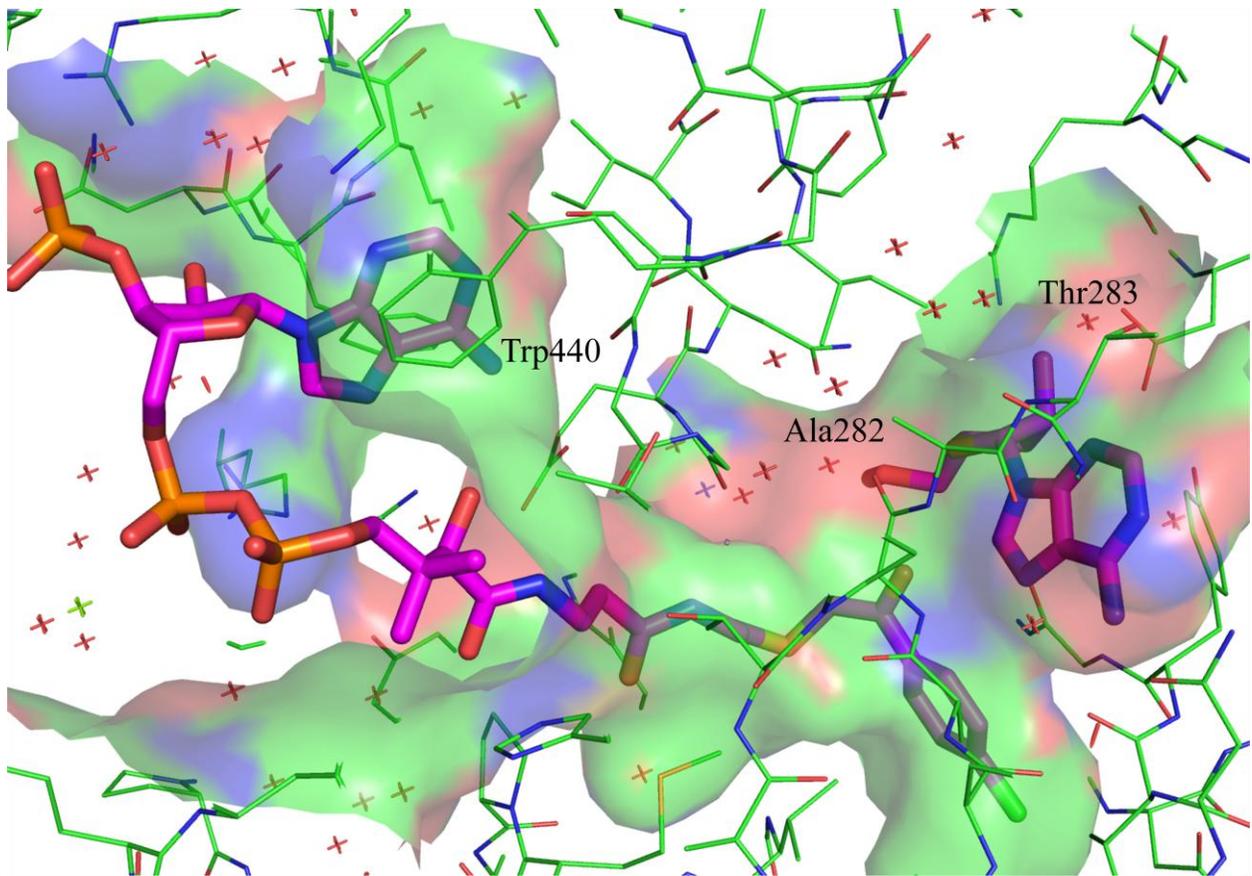


Figure II.3 (cont'd)

The domain movements as they are currently understood for ACSM2A and CBL are as follows. Initially the enzyme is in either the adenylation conformation or the thioesterification conformation, with nothing bound to the active site. The reaction requires that a carboxylate binds to the active site, though this event in and of itself is insufficient to trigger a conformational switch as structures exist in both conformations when only the carboxylate is bound. For example, when 4-chlorobenzoate is bound to CBL (PDB ID 1T5D), CBL is in the adenylation conformation compared to ACSM2A when ibuprofen is bound (PDB ID 2WD9) in which case the thioesterification conformation exists. The binding of the carboxylate can be either before or after ATP binding. However, once ATP binds the enzyme is switched to the adenylation conformation (PDB ID 3CW8 for CBL, PDB ID 3C5E for ACSM2A). At this time the carboxylate becomes adenylated, producing the acyl-adenylate and pyrophosphate. Upon the release of pyrophosphate (PPi) the enzyme returns to the thioesterification conformation (PDB ID 3CW9 for CBL, PDB ID 3B7W for ACSM2A). Switching between the two conformations requires an approximately 140° C-terminal domain rotation which reconfigures the active site for CoA binding and subsequently the proceeding thioesterification reaction (11, 12). Upon creation of the thioester, the product is ejected and the enzyme is now able to switch freely between the two conformations.

It is possible this domain shift weighs on the enzyme's ability to either accept or reject mono-substituted benzoic acids. As the domains move, the surface to which the carboxylate interacts changes dramatically. These changes must be taken into consideration when considering substrate specificities and when designing promiscuous mutants.

A closer look at the active site reveals which residues are directly involved in catalysis in the different conformations. Comparison of the active site residues of BadA and CBL are given in **Section II.3.1**. The proposed mechanism in CBL (which also applies to ACSM2A but with a different substrate) is as follows (**Figure II.4**). First, while in the acyl-adenylate forming conformation 4-chlorobenzoate reacts with ATP to produce the acyl-adenylate. The 4-chlorobenzoate is held in position by a hydrophobic binding pocket comprising of the residues Phe236, Ala237, Tyr238, Ile332, Ser334, Thr335, His339 and Ile340 (**Figure II.3**). This pocket is open towards the back to accommodate the 4-chloro substitution on the aromatic ring. The carboxylate is held in position by hydrogen bonding interactions with Lys492. Adjacent to the carboxylate binding pocket is the ATP binding pocket. ATP is held in place by Phe236, Thr335, Ser191, Gly331, Asp330, Asp412, Arg427 and Gly309. Asp429 connects the ATP binding site with a series of residues (Tyr438, Arg427 and Asp412) which are located in the hinge region between the N- and C-terminal domains. Upon reaction of ATP with 4-chlorobenzoate, diphosphate is released. The release of diphosphate triggers conformational changes in Asp429 and the other hinge region residues which result in the conformational switch from the acyl-adenylate conformation to the thioesterification conformation. The acyl-adenylate is now in position for the reaction with CoA. CoA enters the enzyme through a channel located opposite the ATP binding channel and the carboxylate binding channel (**Figure II.3**) where it is held in place by residues His207, Gly408, Gly409, Pro204, Thr251, His254, Ser407, Lys477, Arg475 and Arg87 with the adenosine group positioned just outside the enzyme. The thiol of CoA reacts with the phosphate group of the acyl-adenylate, producing AMP along with the desired acyl-CoA product. Studies of ACSM2A suggest that the acyl-adenylate is released from the enzyme first, followed by AMP while the enzyme remains in the thioesterification conformation.

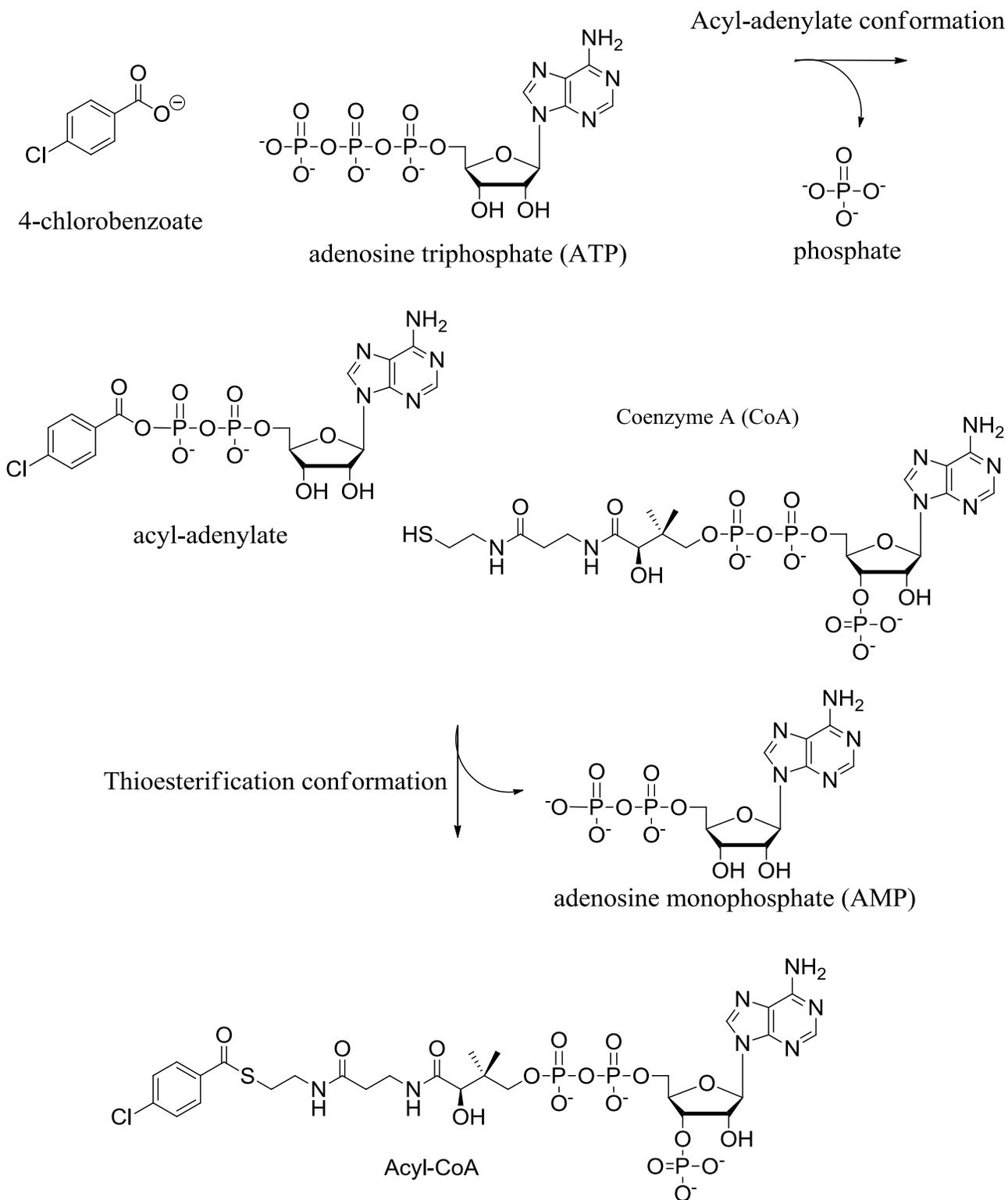


Figure I.4 The proposed mechanism of CoA acylation in CBL includes conformational changes.

To date there is one structure of a Benzoate-Coenzyme A Ligase (BCL), the *Burkholderia xenovorans* LB400 Benzoate-Coenzyme A Ligase from the *box_M* pathway (BCL_M) (PDB ID 2V7B) (13). It has a percent identity of 61% to BadA with several conserved regions between the two. BCL_M is in the adenylation conformation. The structure of BadA might provide insight into the thioesterification conformation for benzoate Coenzyme A ligases.

Unlike BadA which is somewhat promiscuous to substrates in addition to benzoate, BCL_M is highly specific to benzoate with 2-aminobenzoate having the closest relative specific activity at just 12.7 percent. It has no detectible activity for non-aromatic acids and minimal activity for fluorobenzenes (13). This specificity is attributed to a well defined hydrophobic benzoate binding pocket (**Figure II.5**). This pocket is defined by several residues that include Phe236, Ala237, Tyr238, Ile332, Gly333, Ser334, Thr335, His339 and Ile340. Lys520 orients the benzoate via two hydrogen bonds. This architecture is similar to closely related enzymes, leading the authors' to conclude that the second shell residues of the protein are responsible for variations in substrate specificity. For example, in our enzyme the residues mentioned above correspond to Phe226, Ala227, Tyr228, Ile336, Gly327, Ser328, Thr329, His333 and Ile334 respectively. The high conservation between the two yet variation in substrate specificity lends to the authors' conclusion (for more information on sequence similarity see **Table II.5**). Knowing the structure of BadA might improve the understanding of how the architecture of the active site contributes to substrate specificity in the context of these similarities.

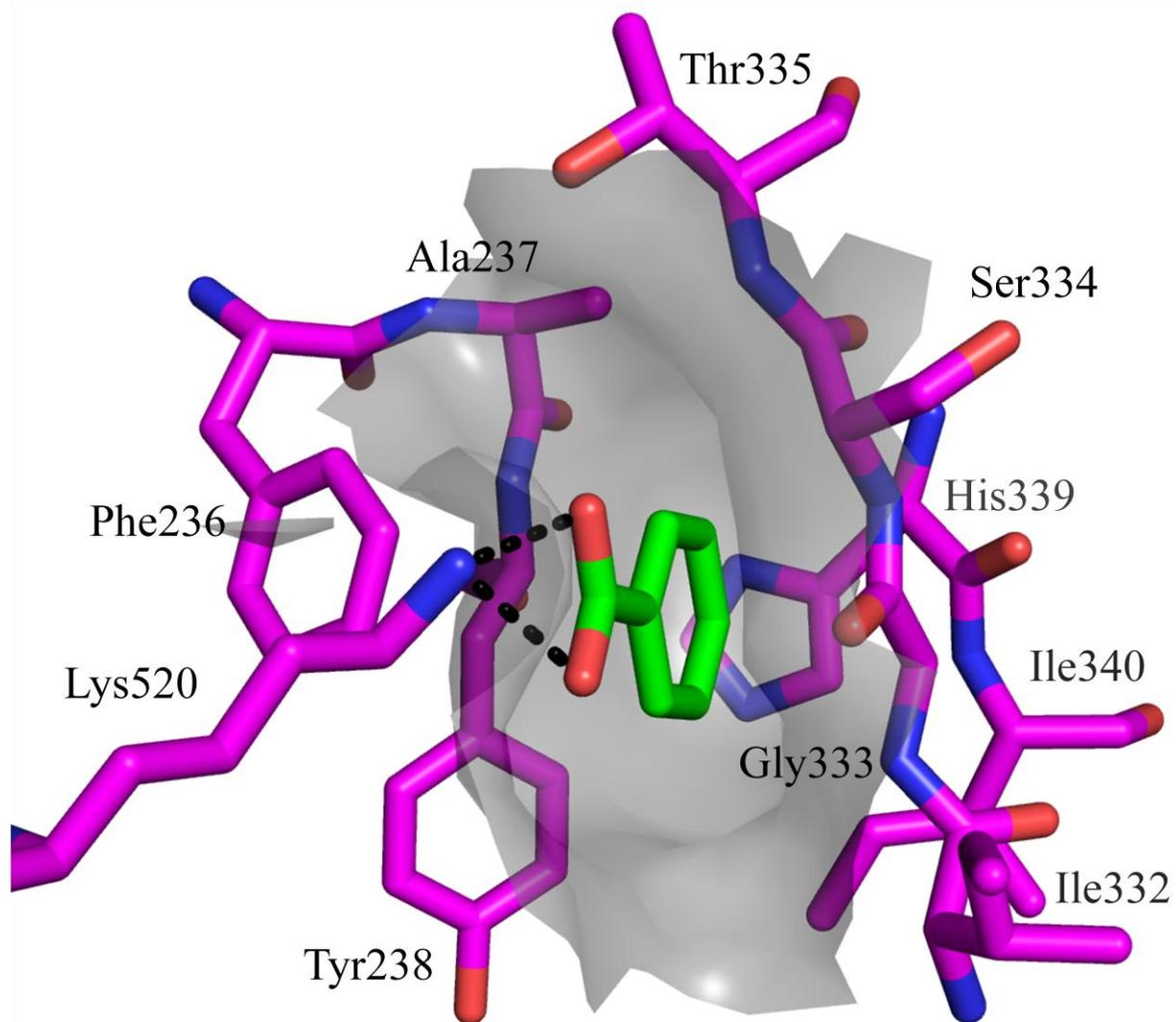


Figure II.5 Key residues in the benzoate binding pocket (magenta) of BCL in the adenylation conformation pack tightly against the benzoate aromatic ring (green) as evident by the surface of the binding site residues which is shown as a grey mesh.

II.2. Experimental Procedures

II.2.1. Crystallization of *Rhodopseudomonas palustris* Benzoate-Coenzyme A

Ligase

Protein for crystallization was provided by Chelsea Thornburg of the Walker Lab. The procedure they used is as follows: the BadA cDNA obtained from *Rhodopseudomonas palustris* was generously provided to the Walker Lab by Caroline Harwood (University of Washington). The gene was amplified by PCR with the following forward primers 5'-TATGAATGCAGCCGCGGTC-3', 5'-TGAATGCAGCCGCGGTCAC-3' (and reverse complements) and subcloned into a pET28a (Novagen) expression vector.

The BadA expression vector was transformed into BL21 (DE3) competent cells (Invitrogen). The cells were grown in LB supplemented with 50 µg/mL kanamycin at 37°C to an optical density of 0.8. Protein production was induced with 0.5 mM IPTG for 5 hours at 18°C. The resulting cell cultures were collected by centrifugation at 6000g. The bacterial pellet was resuspended in buffer (50 mM Na₂PO₄, 300 mM NaCl (pH 8.0)) and protease inhibitor added (EDTA-free Protease Inhibitor Cocktail tablets (Roche)). The suspension was lysed with a Misonix XL 2020 sonicator, clarified at 18,000g for 30 min and the supernatant passed through a 0.2µm filter (Millipore). A Ni²⁺ NTA (Qiagen) column pre-equilibrated with the same buffer as the sample was loaded with the clarified supernatant. The column was washed 5 times with buffer supplemented with 5 % glycerol and 25mM imidazole, and then protein was eluted with 3 column volumes of buffer containing 5 % glycerol and 250mM imidazole.

Fractions containing the enzyme (as evident by SDS-Page) were loaded into a 10kDa cutoff Pierce Dialyzer cassette, and dialyzed overnight against 20mM Tris (pH 8.0) containing 5% glycerol. The protein was next concentrated to approximately 17.5mg/mL (Millipore Amicon

Ultra 30 MWCO). The molecular weight of approximately 65KDa was verified by ESI-MS on a Q-ToF mass spectrometer, and the protein was flash frozen in liquid nitrogen for storage at -80°C.

The Gryphon LCP robot (Art Robbins) was used to screen four different crystallization screens (PEG/pH, Crystal Screen I/II (Hampton Research) Wizard I/II, and Wizard III/VI (Emerald Biosystems)) of 96 conditions a piece using the sitting drop method with plates set both at room temperature. Approximately 45 different conditions produced crystals overnight at room temperature of varying quality; the best being from the Wizard I/II Screen condition A10 (20% (w/v) PEG-2000 MME Tris pH 7.0). Screening of this condition using the hanging-drop diffusion method to 0.1M Tris-HCl, pH 7.0, and 15% PEG 3350 produced crystals appropriate for x-ray data collection (**Figure II.6**).

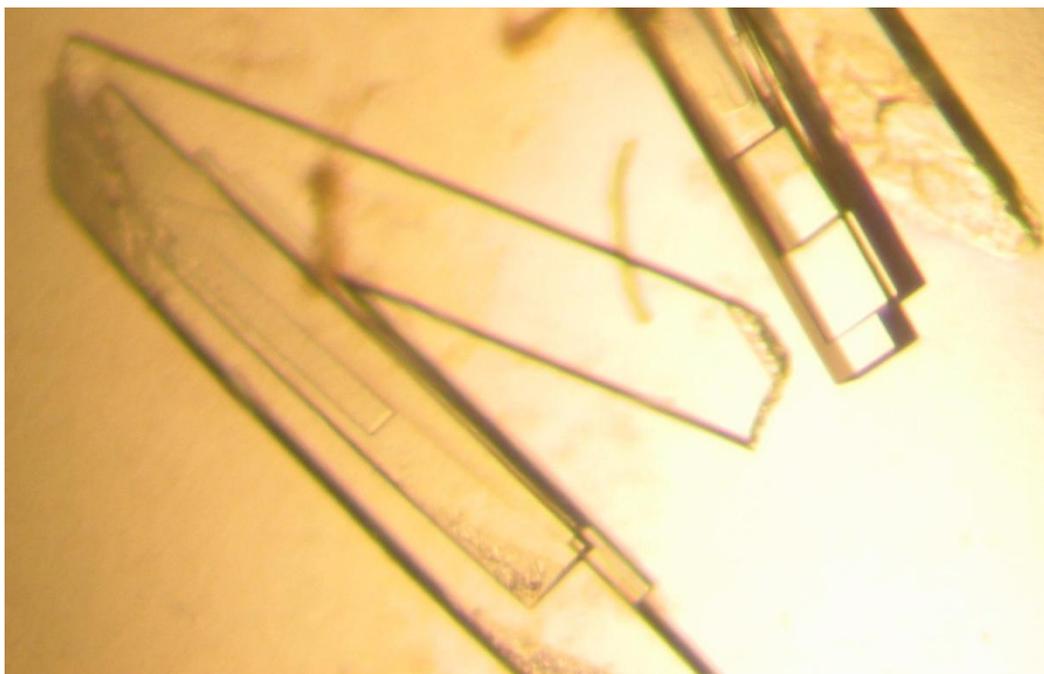


Figure II.6 Crystals of *Rhodospseudomonas palustris* Benzoate-Coenzyme A Ligase

II.2.2. Soaking and Co-crystallization Experiments

Initially, crystals of BadA were soaked for approximately 10 minutes in saturated solutions of various carboxylic acids buffered to pH 7.0. Though no cracking was observed and non-soaked crystals gave excellent diffraction, these crystals when irradiated did not give clear diffraction regardless of the acid used or the amount of time the crystals were soaked.

Therefore, co-crystallization was employed. Protein in 20 mM Tris buffer (pH 8.0) with 5% glycerol at 17.5mg/mL concentration was aliquoted into 25 μ L portions. Various acid solutions were prepared by saturating 1.0 mL of 1M Tris buffer (pH 8.0) (**Figure II.7**). Centrifugation removed the excess acid out of solution. All of the samples were chilled on ice before 5 μ L of one of the different saturated acid solutions were slowly added to the protein solutions. For the co-crystallization of ATP, CoA, and benzoic acid the μ M concentration of the protein was first calculated. CoA (10mM), benzoic acid (100mM) and adenosine triphosphate (ATP) (20mM) were then added in concentrations equimolar to the protein. There were no signs of protein precipitation during ligand addition in all above mentioned cases. After 10 minutes of protein-ligand incubation on ice the crystal screen was set up with all wells initially precipitating. The following day large crystals were observed with the best resulting from a pH range of 6.5-7.5 (20 mM Tris buffer) and a PEG 3350 concentration of 15%.

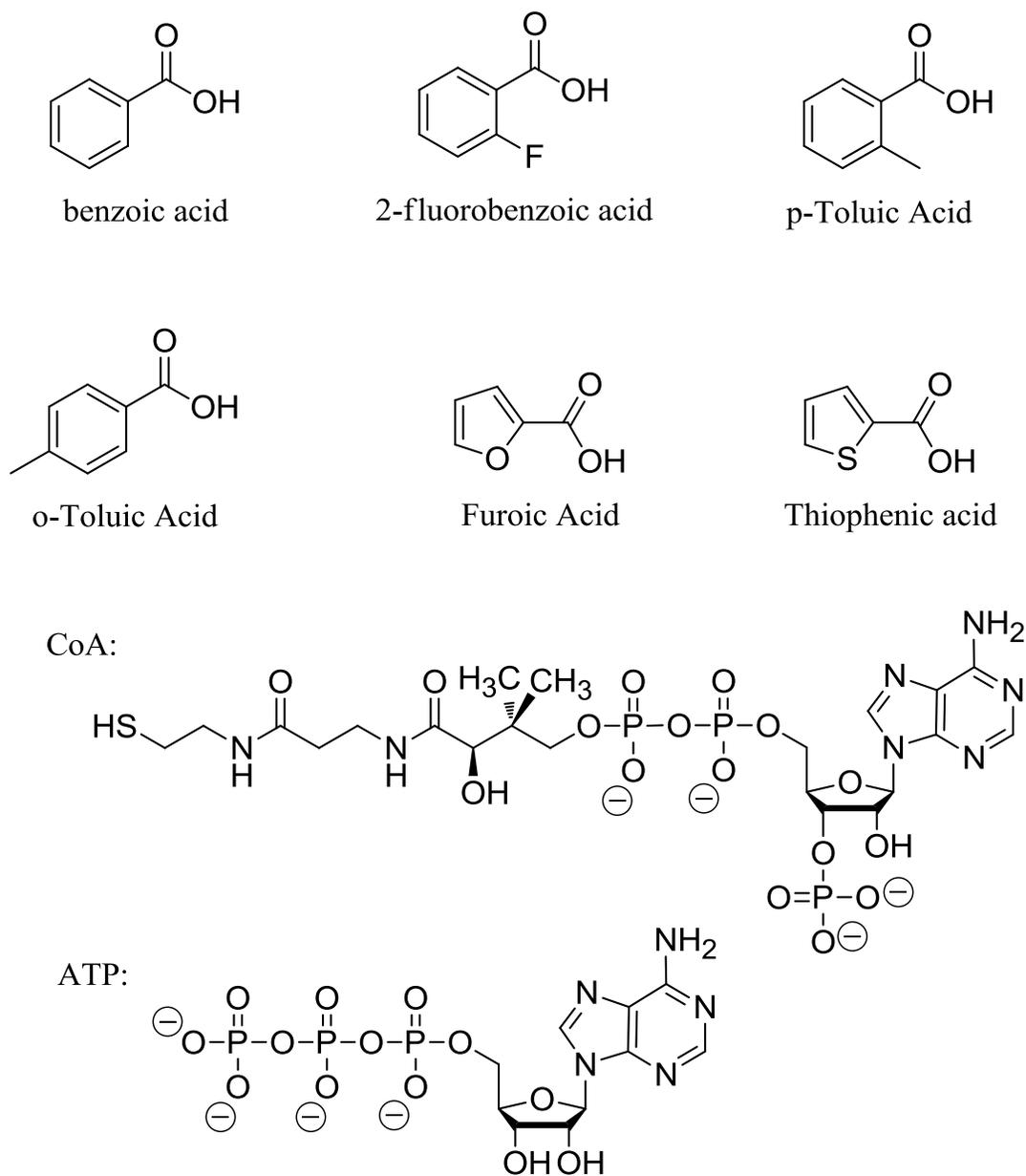


Figure II.7 Ligands used in the co-crystallization experiments of BadA.

Crystals were soaked in cryoprotectant (0.1M Tris-HCl (pH 7.0), 15% PEG 3350, 30% glycerol), mounted in CrypLoops (Hampton Research) and flash frozen in liquid nitrogen. Data was collected at the LS-CAT beamline at Argonne National Laboratories.

II.2.3. Structure Determination

Raw diffraction data was indexed, processed and scaled using the HKL2000 program package (14). A search of the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) revealed one known structure for a benzoate CoA Ligase, a Benzoate-CoA Ligase from *Burkholderia xenovorans* LB400 (PDB ID 2V7B) (13). Swiss-Model (15) produced a threaded homology model based on the PDB model of 2V7B and the known amino acid sequence of BadA. The structure was solved by molecular replacement using this model and the MOLREP program in the CCP4 suite of programs (16) using this model which produced a dimer. REFMAC5 in the CCP4 suite (16) produced the initial density maps, with an initial $R_{\text{cryst}}/R_{\text{free}}$ (%) of 0.4417/0.4742 and a correlation coefficient of 0.6534. At this time the N-terminal domain was well positioned within the density, however the C-terminal domain was not. Therefore, the model was corrected using Buccaneer to thread the amino acid sequence into the density. Further corrections were made manually in COOT (**Table II.2.3.1**, **Table II.2.3.2**, **Table II.2.3.3**). Jligand version 1.0.9 (17) was used to generate the necessary ligands.

Table II.1 Data collection and structure refinement statistics for the structures containing benzoic acid, p-toluic acid and 2-fluorobenzoic acid.

Ligand:	benzoic acid	p-toluic acid	2-fluorobenzoic acid
wavelength (Å)	0.97872	0.97872	0.97872
total reflections	705943	434363	337716
unique reflections	93683	84971	83722
space group	P 2 1	P 2 1	P 2 1
unit-cell parameters	$a = 58.652$ $b = 96.018$ $c = 95.376$ $\alpha = \gamma = 90^\circ$ $\beta = 104.648^\circ$	$a = 58.718$ $b = 95.370$ $c = 95.814$ $\alpha = \gamma = 90^\circ$ $\beta = 104.610^\circ$	$a = 58.92$ $b = 95.736$ $c = 98.641$ $\alpha = \gamma = 90^\circ$ $\beta = 110.385^\circ$
molecules per ASU	2	2	2
resolution range (Å)	50.00 – 1.80	50.00 – 1.86	50.00 – 1.87
completeness (%)	98.7 (97.6)	99.0 (95.1)	99.9 (99.2)
I/ σ	21.7 (2.73)	21.61 (2.53)	16.9 (2.65)
R _{merge} (%)	8.9 (63.7)	8.2 (52.2)	10.0 (41.6)
structure refinement			
resolution (Å)	50.00 – 1.80	50.00 – 1.86	50.00 – 1.87
R _{cryst} /R _{free} (%)	0.1533/0.1910	0.1584/0.1940	0.1603/0.2024
rmsd from ideal values			
bond length (Å)	0.0293	0.0303	0.0271
bond angle (deg)	2.2109	2.2124	2.0498
average B factor	19.686	20.665	17.234
PDB #	4EAT		

Table II.2 Data collection and structure refinement statistics for the structures containing o-toluic acid, 2-furoic acid and thiophenic acid.

Ligand:	o-toluic acid	2-furoic acid	thiophenic acid
wavelength (Å)	0.97872	0.97872	0.97872
total reflections	378161	547422	460337
unique reflections	103806	147005	93314
space group	P 2 1	P 2 1	P 2 1
unit-cell parameters	$a = 58.426$ $b = 94.601$ $c = 94.868$ $\alpha = \gamma = 90^\circ$ $\beta = 104.566^\circ$	$a = 58.590$ $b = 95.465$ $c = 95.271$ $\alpha = \gamma = 90^\circ$ $\beta = 104.777^\circ$	$a = 58.799$ $b = 94.821$ $c = 95.727$ $\alpha = \gamma = 90^\circ$ $\beta = 104.872^\circ$
molecules per ASU	2	2	2
resolution range (Å)	50.00 – 1.73	50.00 – 1.54	50.00 – 1.80
completeness (%)	99.7 (98.0)	98.0 (94.6)	99.6 (95.6)
I / σ	20.99 (2.15)	28.93 (3.48)	26.9 (2.8)
R _{merge} (%)	7.5 (39.7)	4.2 (28.6)	9.3 (42.8)
structure refinement			
resolution (Å)	50.00 – 1.73	50.00 – 1.54	50.00 – 1.80
R _{cryst} /R _{free} (%)	0.1582/0.1896	0.1601/0.1962	0.1583/0.1925
rmsd from ideal values			
bond length (Å)	0.0292	0.0297	0.0291
bond angle (deg)	2.3270	0.1765	2.2140
average B factor	19.993	21.754	24.298

Table II.3 Data collection and structure refinement statistics for the structure containing benzoic acid ligated to adenosine monophosphate (AMP).

Ligand: **benzoic acid and AMP**

wavelength (Å)	0.97872
total reflections	382164
unique reflections	92794
space group	P 2 1
unit-cell parameters	$a = 58.559$ $b = 95.422$ $c = 95.684$ $\alpha = \gamma = 90^\circ$ $\beta = 104.513^\circ$
molecules per ASU	2
resolution range (Å)	50.00-1.80
completeness (%)	98.1 (96.1)
I / σ	20.59 (3.24)
R _{merge} (%)	8.3 (35.5)

structure refinement

resolution (Å)	50.00 - 1.80
R _{cryst} /R _{free} (%)	0.1478/0.1837

rmsd from ideal values

bond length (Å)	0.0292
bond angle (deg)	2.2665
average B factor	16.961

II.3. Results and Discussion

II.3.1 Crystal Structures of *Rhodopseudomonas palustris* Benzoate-Coenzyme

A Ligase

The overall domain folds of BadA are typical of this family of ligases. Residues 1-434 constitute the N-terminal domain which contains the core of the hydrophobic active site. Residues 435-522 comprise the C-terminal domain. The two domains are positioned relative to one another in the thioesterification conformation (11). This conformation, hallmarked by a 140° C-terminal domain rotation relative to the N-terminal domain from the adenylation conformation, is found in such structures as the 4-chlorobenzoyl CoA bound 4-chlorobenzoate: CoA ligase (PDB ID 3CW9). The adenylation conformation is found in the benzoate bound Benzoate CoA Ligase (BCL_M) (PDB ID 1T5D) as well as the acyl-adenylate bound 4-chlorobenzoate:CoA ligase (PDB ID 3CW8) (**Figure II.8**).

In the case of CBL, the adenylation conformation is necessary for the reaction of benzoate with adenosine triphosphate (ATP) (PDB ID 3CW8). The enzyme then converts to the thioesterification conformation after the release of pyrophosphate in preparation of the thioesterification of benzoate (12) (PDB ID 3CW9). The benzoate bound structure of BCL_M (PDB ID 2V7B) is in the adenylation conformation, though the pyrophosphate is not observed in the crystal structure. Given the similarity in domain folds BadA should function in a similar way to that of BCL_M and CBL.

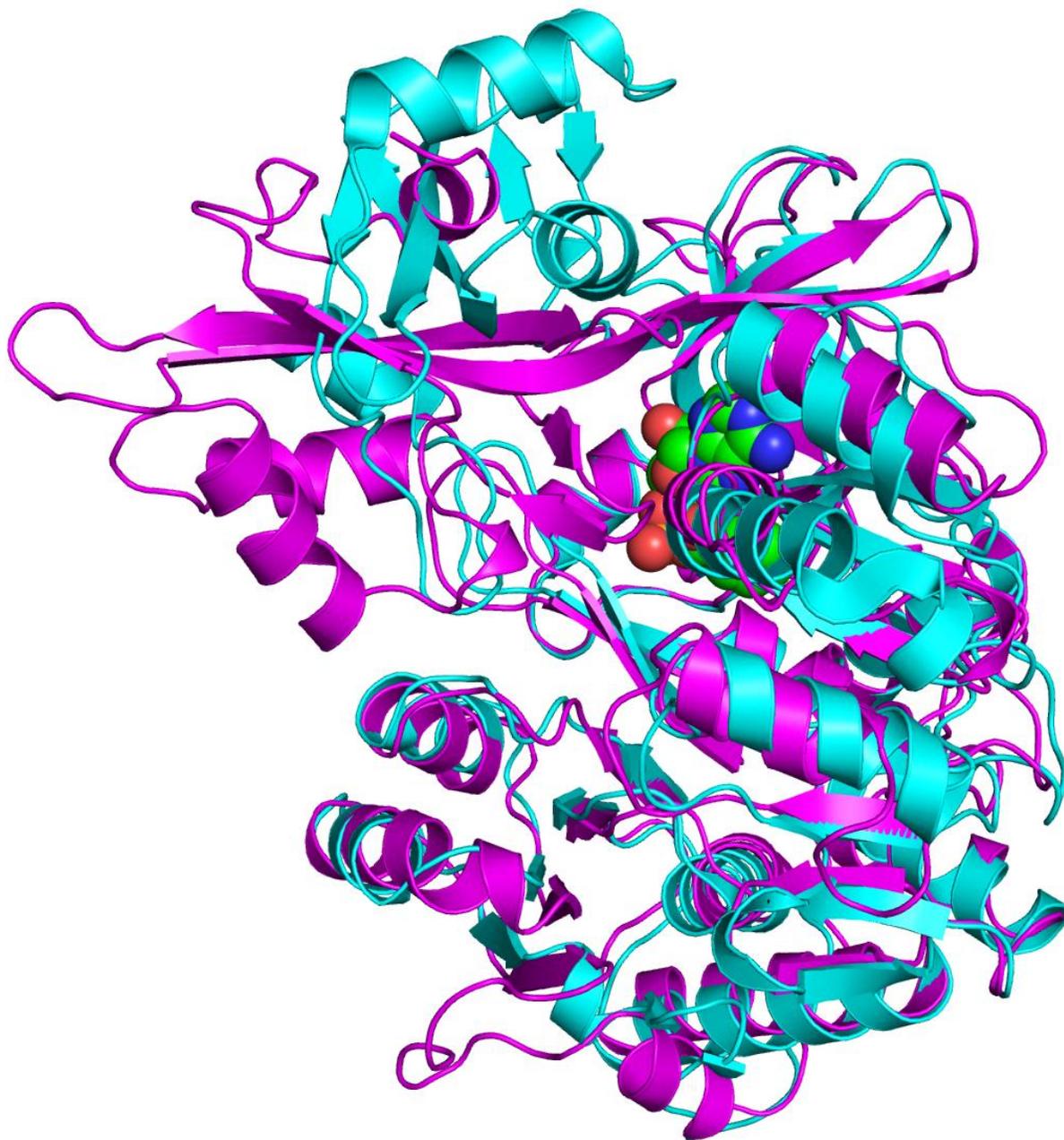


Figure II.8 Overlay of acyl-adenylate bound CBL (cyan) and acyl-adenylate bound BadA (magenta) showing on the left side the large domain shift indicative of the two separate conformations: the adenylation conformation (CBL) and the thioesterification conformation (BadA). As neither benzoic acid CoA ligase is present in both conformations they must be looked at together to gain insight into the various conformations the enzyme adopts during catalysis. The acyl-adenylate of BadA is highlighted in green spheres.

Our structures therefore show steps in the mechanism not yet observed for the benzoate CoA ligases. The acyl-adenylate bound thioesterification conformation of BadA represents the enzyme just before CoA reaction after the release of pyrophosphate. In addition the structures of BadA with benzoic acids bound demonstrates the form of the enzyme before addition of ATP which would trigger the switch to the adenylation conformation. This information is summarized in **Table II.4**. This suggests that BadA requires ATP to interact with the active site to trigger conversion from the thioesterification conformation to the adenylation conformation. A close examination of the residues involved in benzoate, AMP and CoA binding follows to justify the comparison of CBL and BCL_M to BadA. The amino acids involved in these discussions are included in **Table II.5**.

Table II.4 Comparison of different conformations found in ATP dependent CoA ligases.

Protein	PDB ID	Ligand in active site	Conformation
BCL _M	2V7B	benzoate	adenylation
CBL	1T5D	4-chlorobenzoate	adenylation
	3CW8	4-chlorobenzoyl-AMP	adenylation
	3CW9	CoA	thioesterification
ACSM2A	3B7W	apo	thioesterification
	3EQ6	Butyryl Coenzyme A and AMP	thioesterification
	3C5E	ATP	adenylation
	3GPC	CoA	close to adenylation
	3DAY	AMP-CPP non-hydrolysable ATP	adenylation
	2WD9	ibuprofen	thioesterification
	2VZE	AMP	thioesterification
BadA	4EAT	benzoate	thioesterification
		benzoyl-AMP	thioesterification
		2-fluorobenzoate	thioesterification
		furonic acid	thioesterification
		2-toluic acid	thioesterification
		thiophenoic acid	thioesterification

Table II.5 Multiple sequence alignment of BadA, BCL_M (BCLm), CBL and ACSM2A using Clustal 2.1 (18). The conserved A8 domain is highlighted in yellow. The conserved A10 domain is highlighted in blue (19). Residues that are part of the substrate binding pocket are marked in light grey. Residues involved in binding the carboxylate are highlighted in dark grey. Residues involved in AMP binding are highlighted in green. Residues involved in the hinge movement between the N- and C-terminals are marked in red. Underlined residues are involved in CoA binding or are suspected to be involved in CoA binding as evident by structural overlays with CBL where structures of a CoA bound enzyme do not exist (BadA and BCL_M).

BadA	-----MN--AAAVTPPPEKFNFAEHLQQT	22
BCLm	-----MEALLEKAANPPAATVEAPPALFNFAAYLFRL	32
CBL	-----MQTVNEMLRRA	11
ACSM2A	MHWLRKVQGLCTLWGTQMSRRTLYINSRQLVSLQWGHQVEVPAKFNFASDVLHDHWADMEKA	60
	: :	
BadA	NRVRPDKTAFVDDIS----SLSFAQLEAQTRQLAAALR-AIGVKREERVLLMLDGTDPW	77
BCLm	NETRAGKTAYIDDTG----STTYGELEERARRFASALR-TLGVHPEERILLVMLDTVALP	87
CBL	ATRAPDHICALAVPARG--LRLTHAELRARVEAVAARLH-ADGLRPQQRVAVVAPNSADV	68
ACSM2A	GKRLPSPALWVWNGKKGKELMWNFRELSENSQQAANVLSGACGLQRGDRVAVVLPVPEWW	120
 : * . . * * : * : : * : : :	
BadA	VAFLGAIYAGIVPVAVNTLLTADDYAYMLEHSRAQAVLVSGALHPVLKAALTKSDHEVQR	137
BCLm	VAFLGALYAGVVPVAVNTLLTPADYVYMLTHSHARAVIASGALVQNVQALESAEHDGCQ	147
CBL	IAILALHRLGAVPALLNPRKLSAELAELIKRGEMTAAVIA--VGRQVADAIQSGSGARI	126
ACSM2A	LVILGCIRAGLIFMPGTIQMKSTDILYRLQMSKAKAIVAGDEVIQEVDTVASECPSLRIK	180
	: : * . * : : : . . * : . : : . . .	
BadA	VIVSRPAAPLEPGEVDFAEFVGAHAPLEKPAATQADDPAFWLYSSGSTGRPKGVVHTHAN	197
BCLm	LIVSQPRESEPR LAPLFEELIDAAAPA KAAATGCDDIAFWLYSSGSTGKPKGTVHTHAN	207
CBL	I FLGDLVRDGE P-----YSYGPPIEDPQREPAQPAFIFYTSGTTGLPKAAIIPQR-	176
ACSM2A	LLVSEKSCDGLWN---FKKLLNEASTTHHCVETGSQEASAIYFTSGTSGLPKMAEHSYSS	237
	: : : : : : : * * : . .	
BadA	PYWTSELYGRNTLHLRE--DDVCFSAAKLFFAYGLGNALTFPMTVGATLLMGERPTDA	255
BCLm	LYWTAELYAKPILGIAE--NDVVFSAAKLFFAYGLGNLTFPLSVGATAILMAERPTADA	265
CBL	AAESRVLFMSTQVGLRHGRHNVLGLMPLYHVVGFFAVLVAALALDGTYYVVVEEFRPVDA	236
ACSM2A	LGLKAKMDAG-WTGLQA--SDIMWTISDTGWILNLC SLMEPWALGACTFVHLLPKFDPL	294
	. : : : : : : . * . : : . .	
BadA	VFKRWLGGVGGVKPTV FYGAPTGYAGMLAAPNLP--SRDQVALRLASSAGEALPAEIGQR	313
BCLm	IFARLVEHR----PTV FYGVPTLYANMLVSPNLP--ARADV AIRICTSAGEALPREIGER	319
CBL	LQLVQQEQV----TSLFATPTHLDALAAAAAHAGSSLKLD SLRHVTFAGATMPDAVLET	291
ACSM2A	VILKTLSSYP---IKSMMGAPIVYR-MLLQQDLS--SYKFPHLQNCVTVGESLLPETLEN	348
	: . : . * : . : : : . * : : :	
BadA	FQRHFGLDIVDGIGSTEMLHIFLSNLPDRVRYGTTGWVPVPGYQIELRGDGGGPVADGEPG	373
BCLm	FTAHFGEIILDGIGSTEMLHIFLSNRAGAVEYGTGTRPVPGYEIELRDEAGHAVPDGEVG	379
CBL	VHQHLPGEKVNIIYGTTEAMNSLYMRQPKTGTEMAPGFFSEVRIVRIGGGVDEIVANGEEG	351
ACSM2A	WRAQTGLDIRESYGQTEGLTCMVSKTMKIKPGYMGTAASCYDVQIIDDKGNVLP PGTEG	408
	: : : * * . * : : * *	

Table II.5 (cont'd)

BadA	DLYIHG----PSSATM-YWGNRAKSRDTFQGGWTKSGDKYVRNDDGSYTYA GRTDDMLKV	428
BCLm	DLYIKG----PSAAVM-YWNNREKSRATFLGEWIRSGDKYCRLPNGCYVYA GRSDDMLKV	434
CBL	ELIVAA----SDSAFVGYLNQPQATAEKLQDGWYRTSDVAVWTPEGTVRIL GRVDDMIIS	407
ACSM2A	DIGIRVKPIRPIGIFSGYVDNPKTAANIRGDFWLLGDRGIKDEDGYFQ FMGRADDIINS	468
	:: : . . * .: : .: . : .* :* ** **::	
BadA	SGIY VSPFEIEATLVQHPGVLEAAVVGVADEHGLTKPKAYVVPR----PGQTLSETELKT	484
BCLm	SGQY VSPVEVEMVLVQHDAVLEAAVVGVDHG-GLVKTRAFVVLKREFAPSEILAE-ELKA	492
CBL	GGENI HPSIEIERVLGTAPGVTEVVVIGLADQRWGQSVTACVVPR----LGETLSADALDT	463
ACSM2A	SGYR IGPSEVENALMEHPAVVETAVISSPDPV <u>R</u> GEVVKAFVVLASQFLSHDPEQLTKELQ	528
	. * : * * : * . * . * * . * . . . * * * :	
BadA	FIKDR-LAPYKYPRSTVFVAELPKTATGKIQRFKLREGVLG-----	524
BCLm	FVKDR-LAPHKYPRDIVFVDDLPKTATGKIQRFKLREQ-----	529
CBL	FCRSSELADFKRPKRYFILDQLPKNALNKVLRRLVQQVSS-----	504
ACSM2A	QHVKSVTAPYKYPRKIEFVLNLPKTVTGTGIQRAKLKRDKEWKMSGKARAQ	577
	. * . * * : : : * * * . . * : * * :	

When the active site of BadA with benzoic acid bound in the thioesterification conformation is compared to that of BCL_M in the acyl-adenylation conformation, we found that several of the residues that are conserved between the two enzymes (including Ala227, Tyr228, Ile336, Gly327, Ser328, Thr329, His333 and Ile334) occupy almost exactly the same locations. This likely explains why BadA has similar specificities to sterically challenged substrates as BCL_M. The only exception is Phe226 (Phe236 in BCL_M) which is turned approximately 72 degrees away from the active site, opening the channel that CoA would need to occupy. This rotation would be sterically hindered in BCL_M by Thr518 which blocks the rotation. Therefore the switch from the adenylation conformation to the thioesterification conformation likely removes Thr518, allowing Phe236 to swing out of the CoA binding channel in BCL_M (**Figure II.9**).

In CBL this space is occupied by His207 which interacts with Val209 while in the adenylation conformation. The switch to the thioesterification conformation introduces a hydrogen bond interaction with Glu410 which moves His207 out of the CoA binding channel to a similar trajectory seen in BadA for Phe226. Should a structure of BadA in the adenylation conformation become available it would be predicted that Phe226 would be in a conformation similar to that of His207 of CBL and Phe236 of BCL_M. This conformation would block the CoA channel until benzoate and ATP react.

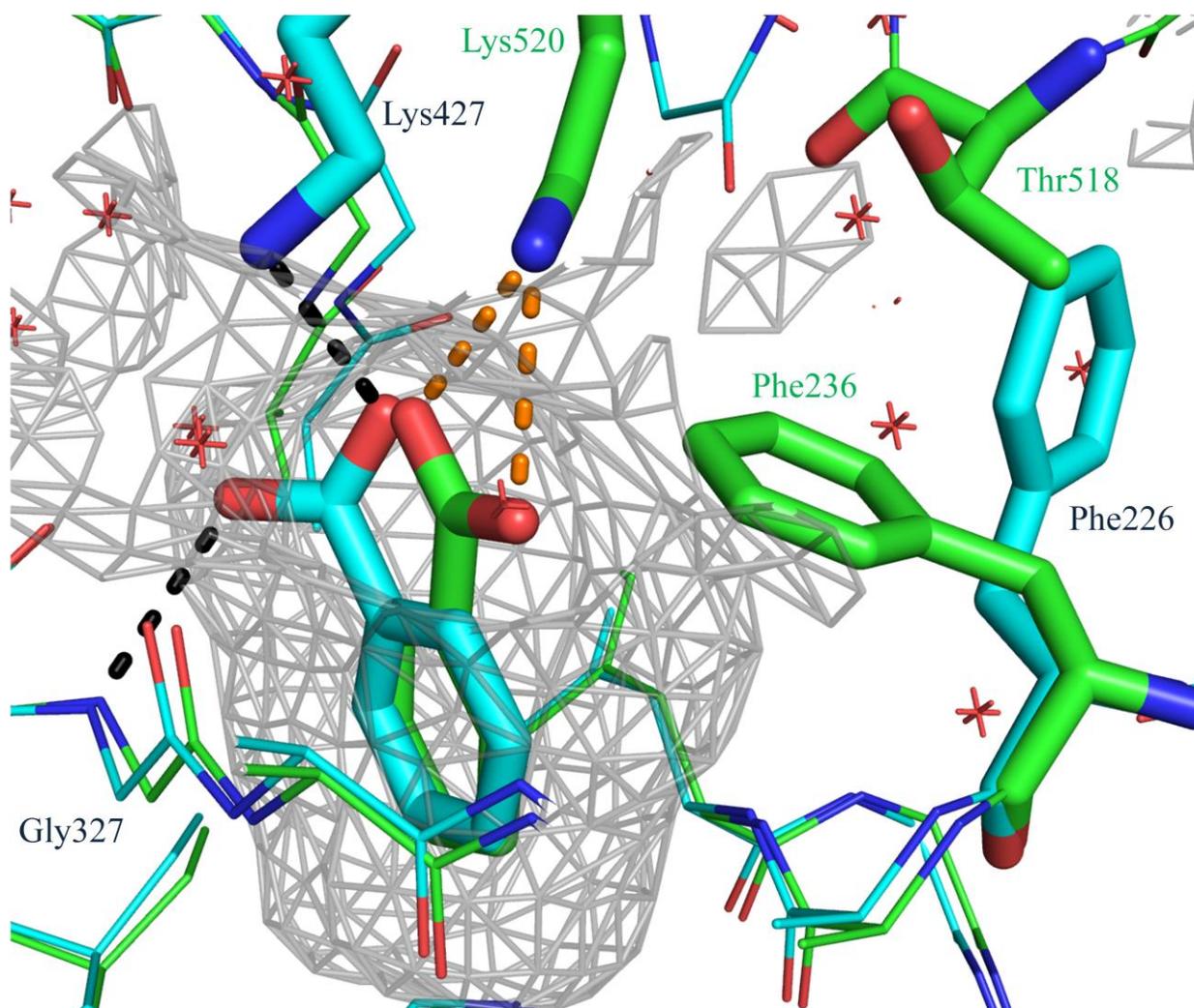


Figure II.9 The switch from the adenylation conformation to the thioesterification conformation in BadA (cyan) allows Phe226 to swing out of the CoA binding channel. Residues in BCL_M are marked in green. Hydrogen bonds between the carboxylates and close residues are marked in black for BadA and orange for BCL_M. Grey mesh representing the interior surface of BCL_M shows that the CoA channel is completely blocked by Phe236.

In BadA the phenyl ring of the substrate benzoate overlays nicely with that of BCL_M except that the carboxylate is twisted out of the plane of the phenyl ring by a torsion angle of approximately 32 degrees and rotated relative to the phenyl ring by approximately 60 degrees (**Figure II.9**). The carboxylate is held in place by a hydrogen bond to Lys427 and an additional hydrogen bond to Gly327 (Gly333 in BCL_M and Gly305 in CBL) which allows for this twist. This prevents the carboxylate from being in resonance with the phenyl ring, activating the carboxylate for nucleophilic attack of ATP. The conserved glycine residue likely functions in a similar manner for BCL_M and CBL forming hydrogen bonding interactions to the carboxylate of the benzoates once the enzyme has entered the thioesterification conformation.

This arrangement within the active site appears reminiscent of that for BCL_M where Lys520 holds the carboxylate of benzoic acid via a double handed hydrogen bond. However, Lys427 relates sequentially to Lys433 of BCL_M. They are part of the highly conserved A8 domain. Lys512 is analogous to Lys 433 of BCL_M both of which are part of the conserved A10 domain (19). In BadA, Lys512 is solvent exposed, disordered, and far from the active site due to the N-terminal domain rotation of the thioesterification conformation. Likewise Lys433 of BCL_M is also solvent exposed, disordered, and far from the active site since the enzyme is in the adenylation conformation.

Since the active site architecture structure of BCL_M in the adenylate-forming conformation is similar to that of BadA in the thioesterification conformation it is tempting to conjecture that BadA does not require a C-terminal domain shift to achieve a functioning active site with an open CoA binding channel. This is further supported when considering the structure of the benzyl-AMP bound BadA structure which will be discussed below. Such a difference might explain the difference in substrate recognition. Without the necessity of a large domain movement, more energy would be available for reaction which might allow for less than ideal substrates to react within the hydrophobic binding pocket. However, it has been shown *in vitro* that when Lys512 is acylated or mutated to an alanine BadA becomes inactive (20). If indeed Lys512 is needed for benzoate binding then BadA likely adopts a conformation similar to that of BCL_M adenylation conformation just before adenylation, reverting to the thioesterification conformation upon the release of pyrophosphate. It is likely this conformational change is necessary for ATP to enter the active site.

The binding of benzoic acid derivatives is similar to benzoic acid (**Figure II.10**). BadA adopts the thioesterification conformation and holds the carboxylate as described above. However, in certain cases Lys427 is further away, interacting with the carboxylate through a water mediated interaction. The twist between the carboxylate and the phenyl ring still exists. Those acids containing a substitution at the second position orient such that the addition is pointed towards the opening held by the carboxylate of benzoic acid observed in “conformation 1” of BCL_M showing how these acids are able to bind within the active site. This preference is highlighted by the 4-toluic acid co-crystallized structure in which 2-toluic acid was discovered in the active site despite the acid being 99.8% pure 4-toluic acid.

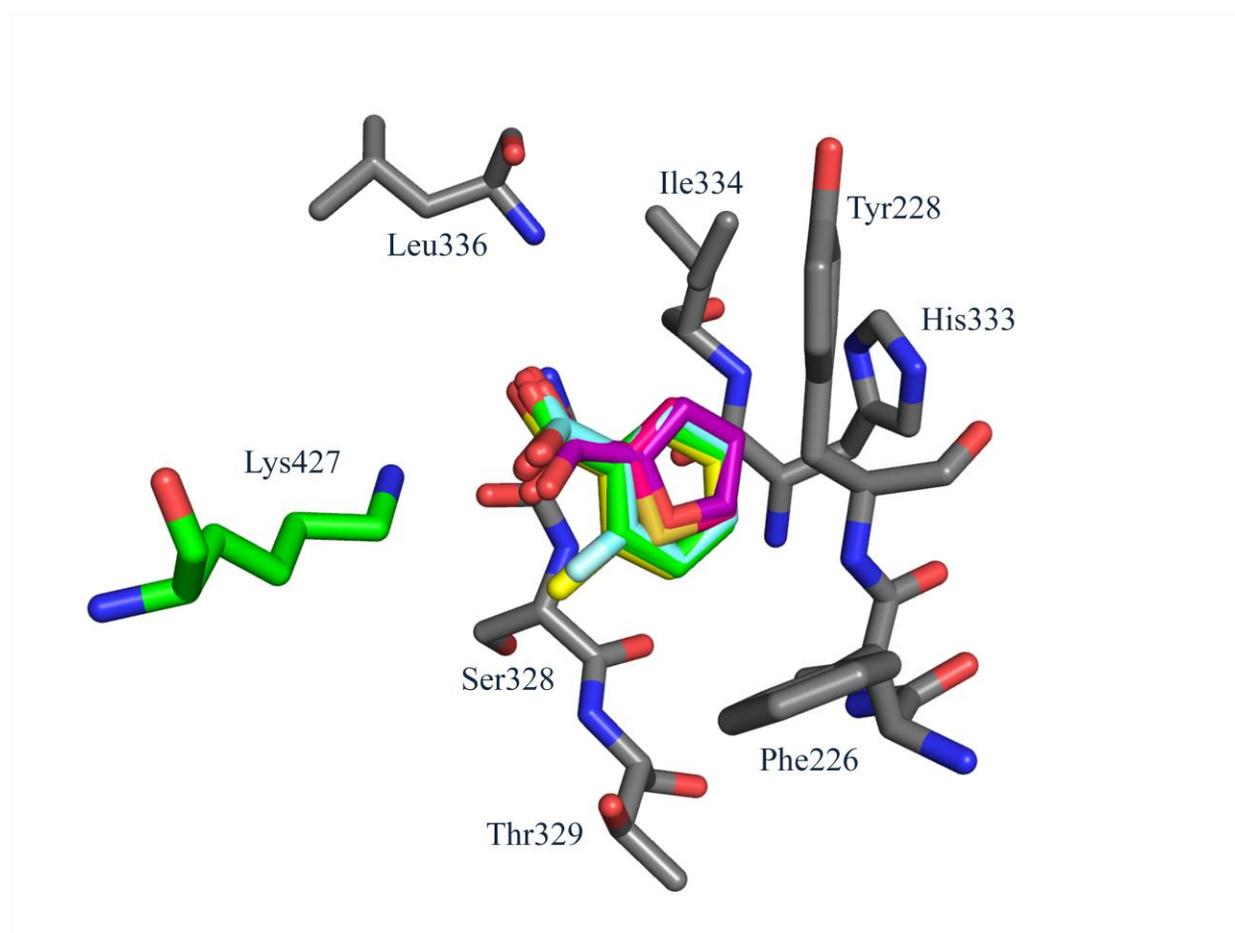


Figure II.10 Overlay of benzoic acid (green) within the active site with 2-fluorobenzoic acid (aqua), 2-toluic acid (yellow), 2-furonic acid (purple) and 2-thiophenic acid (pink).

Benzoic acids substituted at the fourth position conceivably would collide with His333. This collision is avoided in CBL since Met310 occupies this space, giving the 4-chlorobenzoate ligand a polar area to interact with. Mutations at position 333 to smaller or charged side chains might allow for greater acceptance of para substituted benzoic acids. Benzoic acids substituted at the third position would collide with the main chain of amino acids containing Ser328. Therefore altering the amino acid side chains through mutagenesis would likely have little effect on the ability of the enzyme to accept alternate substrates as such changes would not change the main chain trajectory.

The active site architecture of BadA when benzoyl-AMP is bound is different than in the case of CBL. The ligands do not overlap perfectly and interact with different amino acid side chains. Lys427, which originally held the carboxylate of the benzoic acid, now binds to the acyl-adenylate in four locations including two interactions with the alpha-phosphorous group of the AMP, the oxygen of the adenosine ring, as well as the carboxylate oxygen of the benzoyl group (**Figure II.11**). In CBL the interactions are strikingly different due to the altered conformation of the C-terminal region which displaces the corresponding Lys492 21.5Å away to a solvent exposed location. Instead His207 interacts with the bridging oxygen between the phosphorus and the benzoyl. Recall that the corresponding BadA residue Phe226 has no such interaction. Rather the Phe226 is positioned away from the active site reminiscent of the trajectory occupied by CBL His207 after CoA has entered the active site. In fact, this movement of CBL His207 is sterically necessary for CoA to bind. Therefore the position of BadA Phe226 away from AMP allows room for CoA binding.

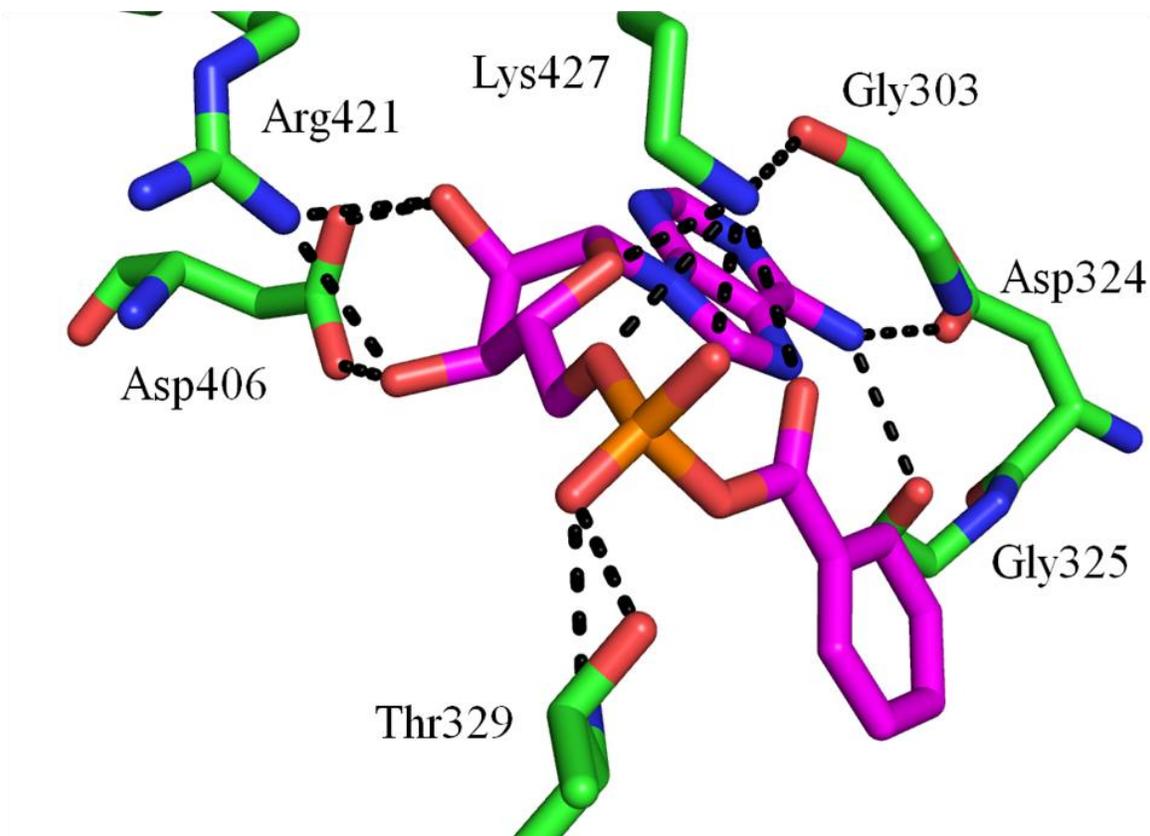


Figure II.11 The binding of benzoyl-AMP in BadA is unique among the known ligase structures. Residues (green) involved in stabilizing the benzoyl-AMP intermediate (magenta) in BadA are shown. Black hash marks denote hydrogen bonding interactions between the ligand and enzyme.

In CBL, Thr307 (Thr329 in BadA) and Thr161 (Ser181 in BadA) interact with the alpha-phosphorus. In BadA, Ser181 is too far from the adenosine to interact with the phosphate while Thr329 instead interacts with carboxylate of the bound benzoic acid. Despite these differences in binding, the acyl-adenylate is positioned in similar conformations. The pucker on the ribose ring is the same for both BadA and CBL. The major difference between the positions of the acyl-adenylate in BadA and CBL is in the torsion angle around C5'. This bond is twisted approximately 78 degrees which brings the O5' oxygen of the phosphate within hydrogen bond distance of Lys427. It also changes the hydrogen bonding pattern observed in BadA for the hydroxyls on the ribose ring, with O3' no longer forming the hydrogen bonds observed in CBL.

In BadA the position of Arg421, Asp406, Gly303, Asp324 and Gly325 make hydrogen bond interactions with the adenylate that are analogous to those found in BCL_M (Arg400, Asp385, Gly281, Asn302 and Ile307 respectively) (**Figure II.12**). Asp324 interacts with the adenosine base N6 amino group and Asp406 and Arg421 interact with the hydroxyl group of the ribose ring

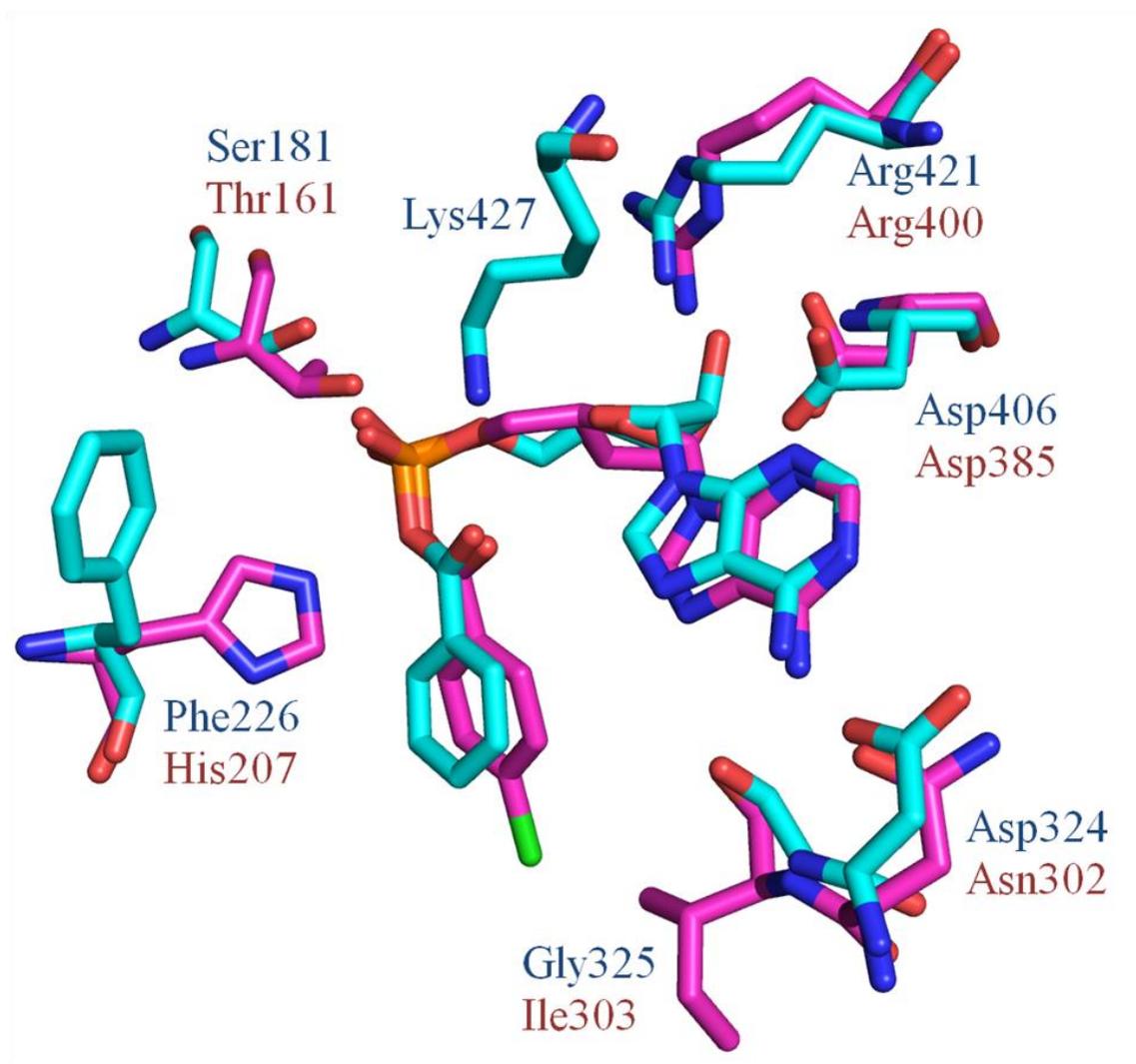


Figure II.12 Overlay of residues involved in benzoyl-AMP binding in BadA (cyan) and CBL (magenta) highlight similarities in the arrangements of the residues with the exceptions of Phe226 and Lys427 in BadA. Residues listed in cyan are those of BadA and residues listed in magenta (typically appearing below those of BadA) are those of CBL.

Though residues Ser192-Gly196 are disordered in BCL_M and other ligases they are well ordered in our structure forming what has been described as a P-loop which is part of the TSG(S/T)-TGxPKG motif (21). This loop interacts with the beta and gamma phosphate groups of ATP in myosin and other P-loop proteins. In the human medium-chain acyl-coenzyme A acyl-CoA synthase ACSM2A ATP bound structure (3C5E), which is in the adenylation conformation, the P-loop is involved in forming the “pyrophosphate pocket” the ATP phosphate groups occupy (10). However there is no evidence in our structure that this loop interacts with the AMP-benzoic acid substrate (**Figure II.13**). Instead the pyrophosphate is no longer bound and the enzyme has switched to the thioesterification conformation.

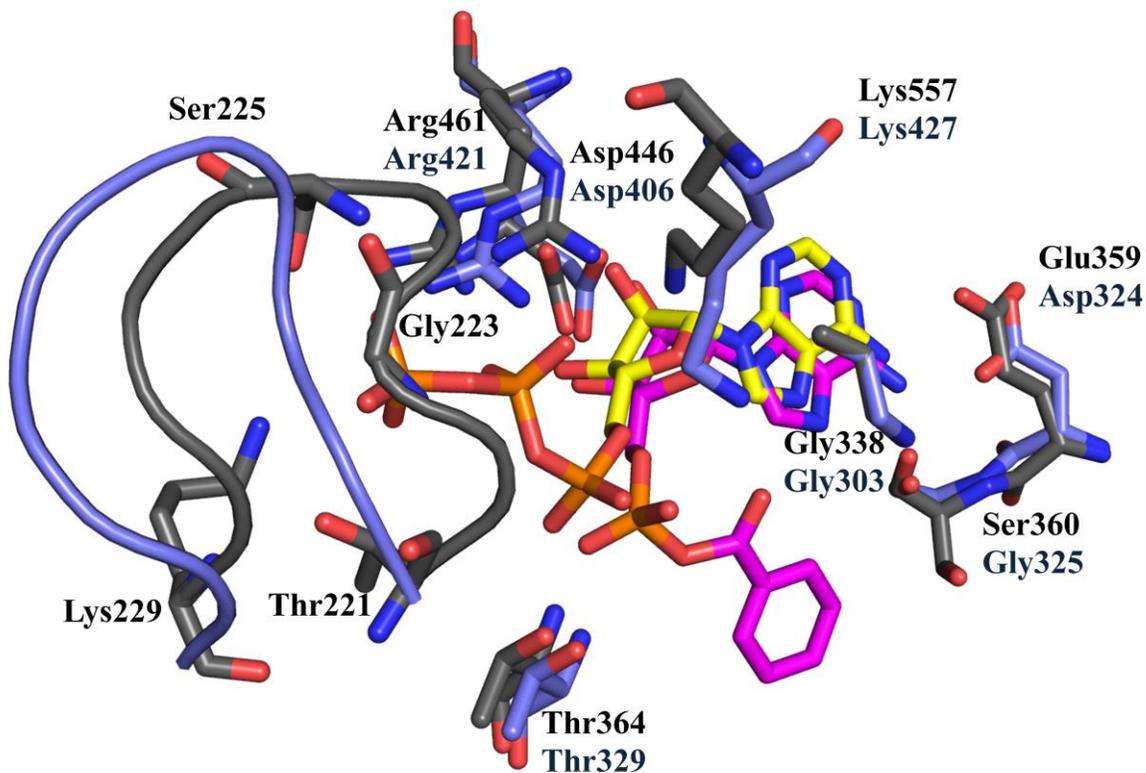


Figure II.13 Comparison of the residues (blue) involved in the acyl-adenylate of BadA (magenta) and the residues (grey) involved in binding ATP (yellow) in Human Medium-chain Acyl-coenzyme A Synthetase ACSM2A. The pocket occupied by the phosphates of ATP in Human Medium-chain Acyl-coenzyme A Synthetase ACSM2A is defined by the P-loop (grey cartoon).

In CBL residues 401- 403 form a type III β -turn separating the N- and C-terminal domains referred to as the hinge region. Asp402 has the greatest change in torsion angle when going from the adenylation conformation to the thioesterification conformation (22). In the adenylation conformation, Arg400 interacts with Asp403. After rotation to the thioesterification conformation Asp402 ion pairs with Arg400 which interacts with the AMP phosphate. When Asp402 is mutated to D402P CoA turnover goes down x1500 fold. The D402A mutant has a CoA turnover decrease x100 fold.

The conformational change from the adenylation conformation to the thioesterification conformation moves Lys492 in CBL away from the active site while introducing Glu410 which repositions His207 out of the CoA binding channel. A loop of Asp280 to Thr283 moves near to ATP. In BadA the corresponding hinge amino acid is Asp423 and it does indeed make hydrogen bond contacts with Arg421 (**Figure II.14**). This again suggests that our AMP bound structure represents a step between the adenylation conformation and the thioesterification conformation where the AMP is primed to react with CoA. However, BadA does not contain His207, but rather Phe226. Therefore the movement of Phe226 cannot be effected by hydrogen bonding interactions as seen in CBL.

The residues involved in this interaction in BadA are however conserved in BCL_M with Tyr438 (Tyr432 in BadA) being just before a disordered loop in the adenylation conformation. Recall that in BCL_M, Thr518 blocks Phe236 (Phe226 in BadA) from rotating out of the CoA binding pocket while the protein is in the adenylation conformation. Therefore, unlike CBL which uses hydrogen bonding interactions to block the channel, BCL_M and Bad A use steric interactions. Knowledge of this interaction may be useful in creating promiscuous mutants of BadA.

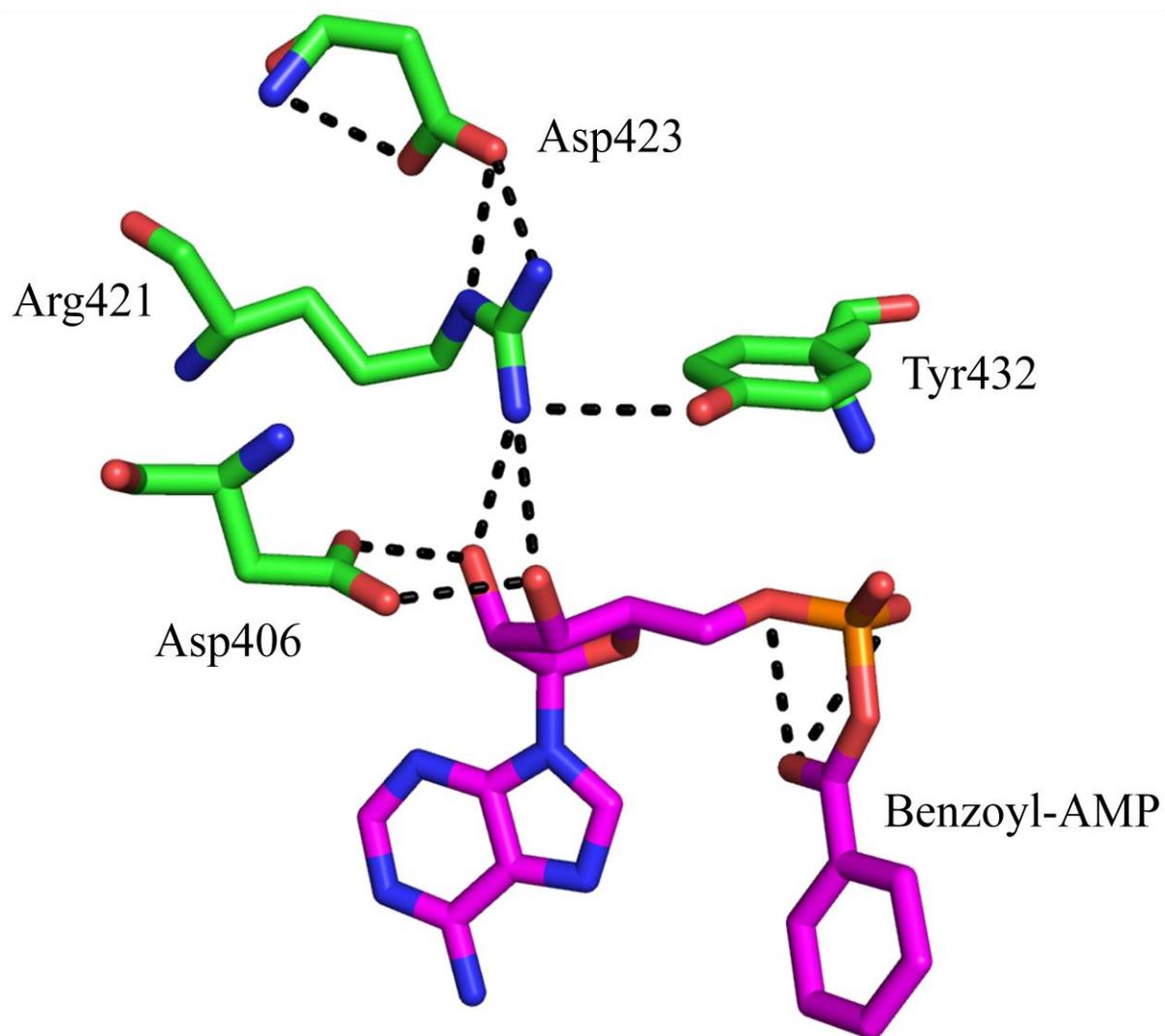


Figure II.14 Interactions between the hinge residues of the N- and C-terminal domain of BadA (green) with benzoyl-AMP bound in the active site (magenta) demonstrates the network of hydrogen bonds formed while BadA is in the thioesterification conformation. Black hash marks denote hydrogen bonds.

Though no structures of a CoA bound BadA are available at this time, we can use the structure of CBL bound to CoA to predict a binding pattern by superimposing PDB ID 3CW9 on the benzoyl-AMP bound structure of BadA (**Figure II.15**). The benzoyl-AMP ligand overlaps with the 4-chlorobenzoyl section of the 4-chlorobenzoyl CoA. The remainder of the CoA chain has no collisions with the enzyme showing the openness of the channel. Where Gly408 and Gly409 interact with the CoA in CBL there are instead close contacts to Ser429 and Gly430 respectively in BadA. Outside the binding channel BadA residues Arg497, Lys487 and Arg250 occupy the locations where Lys477, Arg475 and Arg87 respectively are found, all of which interact with the adenylate of CoA in CBL. These three residues are conserved between BadA and BCL_M suggesting they interact with the adenylate of CoA in a similar manner.

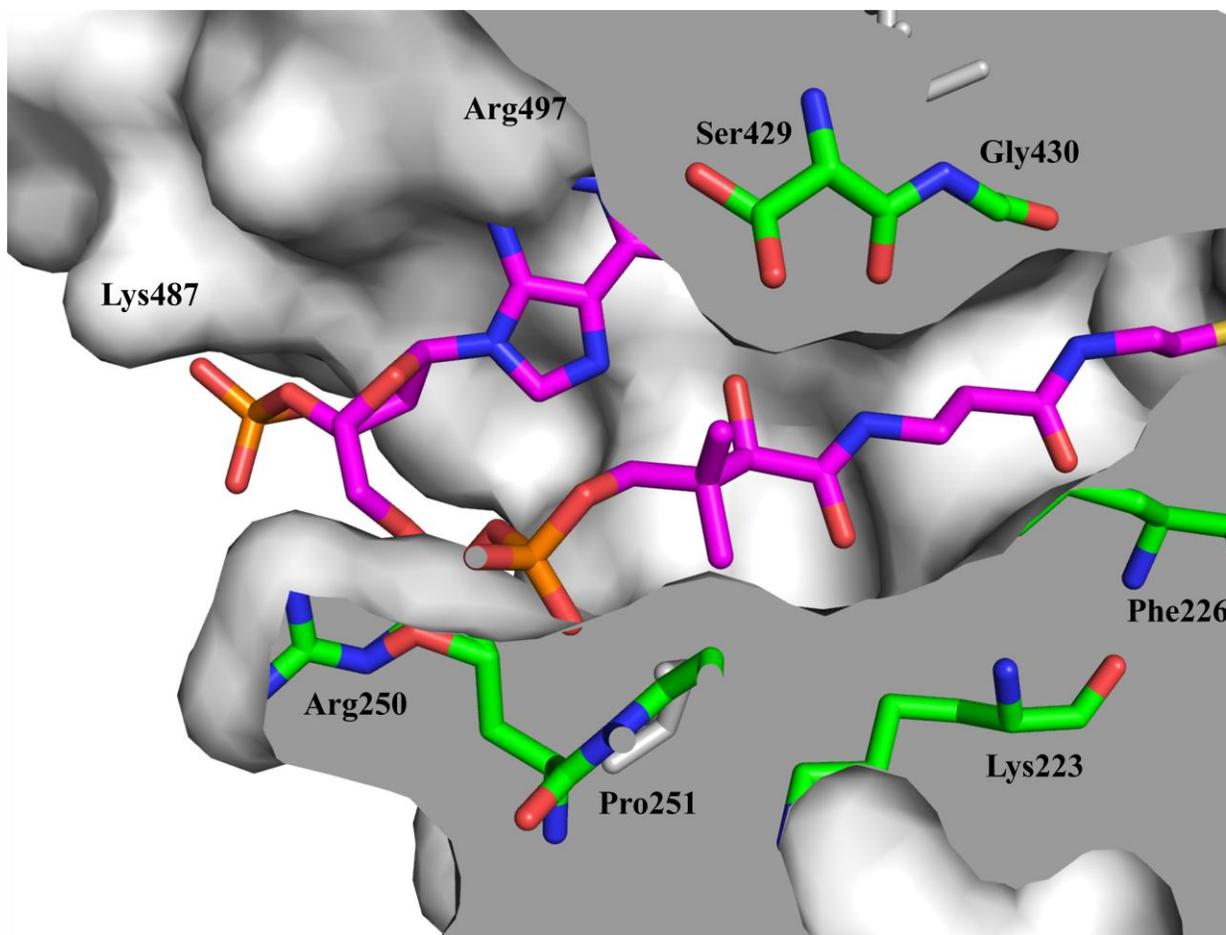


Figure II.15 Overlay of benzyl CoA found in CBL (magenta) with BadA suggests amino acids relevant for CoA binding (green). Rotation of Arg250 would allow for interaction with the CoA phosphate group.

In some of the structures of BadA co-crystallized with benzoic acid derivatives these acids appear in the CoA binding channel as well as the active site. This suggests that part of the reason for benzoic acid inhibition of enzyme activity is the ability of excess benzoic acid to bind to the CoA channel, preventing further reaction.

A mechanism for the conversion of benzoate to benzoyl-CoA has been proposed (13). In it the benzoate, stabilized by His207, nucleophilically attacks ATP at the alpha phosphate, releasing pyrophosphate. However, there is no His207 in our structure. Rather Lys427 is within hydrogen bonding distance of the carboxylate. Presumably it is Lys427 along with Gly327 that

stabilizes the benzoate while twisting the carboxylate out of the plane of the phenyl ring as described above. The ATP phosphates are predicted to be stabilized during this transformation by Thr307 and Thr161 in CBL based on kinetic studies (12). Again, the difference in the binding of the AMP in BadA suggests that Lys427 and Thr329 would stabilize the phosphates during the nucleophilic attack. Again, the residues mentioned in BadA are conserved in BCL_M and thus they likely follow mechanisms similar to one another and unique to CBL. This alteration in ATP binding might be necessary in CBL to facilitate the substitution at the fourth position of the benzoate. Mutations of Lys427 to Ile in BadA might alter the binding of ATP to that observed in CBL thus increasing its promiscuity for para substituted benzoic acid derivatives.

In BadA, initially the carboxylate of the benzoic acids are rotated approximately 60 degrees from where they ultimately end up after adenylation. This rotation towards ATP likely occurs after the conformational change where Lys512 would position the carboxylate for benzoic acid for nucleophilic attack of the alpha phosphate of ATP. The release of pyrophosphate would trigger the return to the thioesterification conformation where Lys427 would now position the acyl-adenylate in such a way that the thio group of the CoA can easily react with the phosphate. In CBL, once the acyladenylate is formed Thr307 and Thr161 stabilize the phosphate of the adenylate for CoA attack. In BadA the interactions of Thr307 and Thr161 would likely be replaced by Lys427 and Thr329. The presence of meta or para substituents would cause collisions with the binding pocket during this initial rotation, and likely explains the enzymes preference for ortho substituents. In the case of ortho substitution there is space for the substituent to rotate towards the ATP binding pocket.

In conclusion, the structures of BadA likely represent different steps in the benzoate transformations then are already available in the PDB for benzoate CoA ligases. The similarity in

the protein architecture and ligands bound, yet difference in the overall tertiary protein structure aids in understanding the mechanism of action. With this understanding it is possible to design mutant constructs that might prove to be more promiscuous than the native protein, allowing for the biosynthetic production of various benzoyl-CoA derivatives for use in biosynthesis.

REFERENCES

REFERENCES

- 1 Schmelz S, Naismith JH (2009). "Adenylate-forming enzymes". *Current Opinion in Structural Biology* **219**:666–671.
- 2 Conti E, Franks NP, Brick P (1996). "Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes". *Structure* **15**: 4(3): 287-98.
- 3 Villemur R (1995). "Coenzyme A ligases involved in anaerobic biodegradation of aromatic compounds". *Can. J. Microbiol.* **41**: 855-861.
- 4 Vetting MW, de Carvalho LPS, Yu M, Hegde SS, Magnet S, Roderick SL, Blanchard JS (2005). "Structure and functions of the GNAT superfamily of acetyltransferases". *Archives of Biochemistry and Biophysics* **433**: 212–226.
- 5 Walker K, Long R, Croteau R (2002). "The final acylation step in Taxol biosynthesis: Cloning of the taxoid C13-side-chain N-benzoyltransferase from *Taxus*". *PNAS.* **99** (14): 9166-9171.
- 6 Gessler JF, Harwood CS, Gibson J (1988). "Purification of benzoate-coenzyme A ligase, a *Rhodoseudomonas palustris* enzyme involved in the anaerobic degradation of benzoate". *Journal of Bacteriology.* **170** (4): 1709 – 1714.
- 7 Kim MK, Harwood CS (1991). "Regulation of benzoate-CoA ligase in *Rhodoseudomonas palustris*". *FEMS Microbiology Letters.* **83**: 199-204.
- 8 Elder DJE, Kelly DJ (1994). "The bacterial degradation of benzoic acid and benzenoid compounds under anaerobic conditions: Unifying trends and new perspectives". *FEMS Microbiology Reviews.* **13**: 441-468.
- 9 Wu R, Reger AS, Cao J, Gulick AM, Dunaway-Mariano D (2007). "Rational Redesign of the 4-Chlorobenzoate Binding Site of 4-Chlorobenzoate: Coenzyme A Ligase for Expanded Substrate Range". *Biochemistry,* **46**: 14487-14499.
- 10 Kochan G, Pilka ES, von Delft F, Oppermann U, Yue WW (2009) "Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A". *J. Mol. Biol.* **388**: 997-1008.
- 11 Reger AS, Wu R, Dunaway-Mariano D, Gulick AM (2008). "Structural Characterization of a 140° domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase". *Biochemistry* **47**: 8016-8025.
- 12 Wu R, Cao J, Lu X, Reger AS, Gulick AM, Dunaway-Mariano D (2008). "Mechanism of 4-chlorobenzoate:coenzyme A ligase catalysis". *Biochemistry* **47**: 8026-8039.

- 13 Bains J, Boulanger MJ (2007). "Biochemical and structural characterization of the paralogous benzoate Coa ligases from *Burkholderia xenovorans LB400*: defining the entry point into the novel benzoate oxidation (box) pathway". *J. Mol. Biol.* **373**: 965-977.
- 14 Otwinowski Z, Minor W (1997). "Processing of X-ray Diffraction Data Collected in Oscillation Mode". *Methods in Enzymology*. **276** (Macromolecular Crystallography, part A): 307-326. C.W. Carter, Jr. & R. M. Sweet, Eds., Academic Press (New York).
- 15 Arnold K, Bordoli L, Kopp J, Schwede T (2006). "The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling". *Bioinformatics* **22**:195-201.
- 16 Collaborative Computational Project, Number 4 (1994). "The CCP4 Suite: Programs for Protein Crystallography". *Acta Cryst.* **D50**: 760-763.
- 17 Lebedev A (2011) "Jligand" Retrieved Fri Jun 3, 2011, from <http://www.ysbl.york.ac.uk/mxstat/JLigand/index.html>
- 18 Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998). "Multiple sequence alignment with Clustal X". *Trends Biochem. Sci.* **23**: 403-405.
- 19 Marahiel MA, Stachelhaus T, Mootz HD (1997). "Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis". *Chem. Rev.* **97**: 2651-2673.
- 20 Crosby HA, Heiniger EK, Harwood CS, Escalante-Semerena JC (2010). "Reversible N ϵ -lysine acetylation regulates the activity of acyl-CoA synthetases involved in anaerobic benzoate catabolism in *Rhodospseudomonas palustris*". *Molecular Microbiology*. **76**(4): 874-888.
- 21 Smith CA, Rayment I (1996). "Active site comparisons highlight structural similarities between myosin and other P-loop proteins". *Biophys. J.* **70**: 1590-1602.
- 22 Wu R, Reger AS, Xuefeng Lu X, Gulick AM, Dunaway-Mariano D (2009). "The Mechanism of Domain Alternation in the Acyl-Adenylate Forming Ligase Superfamily Member 4-Chlorobenzoate: Coenzyme A Ligase". *Biochemistry* **48**: 4115-4125.

Chapter III: SNAPc

III.1. Background

III.1.1. RNAPII and RNAPIII

RNA Polymerase II (RNAP II or Pol II) and RNA Polymerase III (RNAP III or Pol III) transcribe DNA to RNA. Both are multi-protein complexes that recognize pre-initiation complexes (PICs) associated with DNA promoters from which RNA transcription can be initiated (1). Much interest in their structure and function has been generated due to their key role in RNA expression which is the first step in producing both functionally mature non-coding RNAs and protein coding RNAs.

Improper expression of RNA causes many diseases including cancer (2). Understanding how RNAPs directly interact with PICs could highlight weaknesses in the recognition of PICS by RNAP III and ultimately lead to strategies for targeting these diseases through small molecule interactions or gene therapy.

RNAP II consists of 12 subunits in humans as well as in yeast. It is itself incapable of recognizing the promoter without the presence of a PIC (3). It is primarily responsible for the transcription of precursor mRNA, snRNA and microRNA (4, 5). The complete 12 subunit RNA polymerase structure for yeast has been determined to 3.88Å resolution (6) (PDB ID 3FKI).

RNAP III is thought to be primarily responsible for the transcription of “housekeeping” genes producing such non-coding RNAs as ribosomal 5S rRNA and tRNA among other small RNAs (7). It is known to consist of 17 well defined subunits in yeast, (see (7, 8) for reviews). It is also well reviewed in humans (9-11).

III.1.2. Small Nuclear RNA Promoters

A promoter is a specific sequence of DNA either upstream or downstream of the transcription start site that is recognized by transcription factors (12). Some of these factors interact directly with the promoter, recruiting other factors before a polymerase is recruited. The transcription factors that recruit the polymerase are referred to as the pre-initiation complex (PIC). It is the RNA polymerase that ultimately transcribes the DNA to RNA (1).

One such promoter is the U1 small nuclear RNA (snRNA) gene promoter. This promoter contains a Proximal Sequence Element (PSE) that is approximately 55 base pairs (bp) upstream of the transcriptional start site which is recognized by the transcription factor Small Nuclear RNA Activating Protein Complex (SNAPc) (13). The U1 snRNA gene also contains a distal sequence element (DSE) which is approximately 220bp from the transcription start site (14). The DSE is recognized by another transcription factor, Oct1. Together with Oct1 and the transcription factor selenocysteine tRNA-activating factor (Staf) which recognizes the *SphI* postoctamer homology (SPF) sequence (15), SNAPc recruits the TATA Binding Protein (TBP, also referred to as TATA-box Binding Protein), Transcription Factor (TF) IIA, TFIIE and ultimately TFIIIF. These six factors form the PIC needed to recruit human RNA Polymerase II (15).

In comparison, the U6 small nuclear RNA promoter contains the same PSE and DSE with the addition of a TATA Box approximately 25bp downstream of the PSE (17). Though SNAPc and Oct1 are still recruited to the same promoter elements, the TATA Box recruits TBP in the form of Brf-2-TFIIB directly to the DNA. TFIIB consists of TBP, TFIIB Related Factor (BRF2) and B'' (BDP1) (12). SNAPc and TFIIB form a Pol III-specific PIC to which human RNA Polymerase III is recruited (9).

III.1.3. Small Nuclear RNA Activating Protein Complex (SNAPc)

SNAPc is composed of five known subunits: SNAP190, SNAP50, SNAP43, SNAP45 and SNAP19 (18-20) (**Figure III.1**). GST-pull down assays show that SNAP190 is the backbone of the complex, with SNAP43 amino acids 164-268 and the entire sequence of SNAP19 interacting with amino acids 84-133 of the N-terminus of SNAP190. SNAP45 interacts with the C-terminus of SNAP190 from amino acids 1261-1393. SNAP50 interacts tightly with amino acids 1-164 of SNAP43 (9). SNAP190 also contains 41/2 repeats of a Myb like DNA binding domain from amino acids 263 to 503. Immunoprecipitation experiments have shown that this region, along with a zinc finger domain on SNAP50, interact closely with the DNA. From amino acids 888 to 912 is the Oct1 Interacting Region (OIR) (21, 22).

III.1.4. Previous studies of mini-Small Nuclear RNA Activating Complex (mSNAPc)

In order to understand the minimal machinery necessary for snRNA transcription, C-terminal deletion constructs of SNAP190 were produced in which only amino acids 1-505 were expressed. This removed the OIR and the SNAP45 interacting region from the protein. The truncated SNAP190 was expressed as a fusion protein containing an N-terminal glutathione S-transferase (GST) affinity tag with a thrombin cleavable linker.

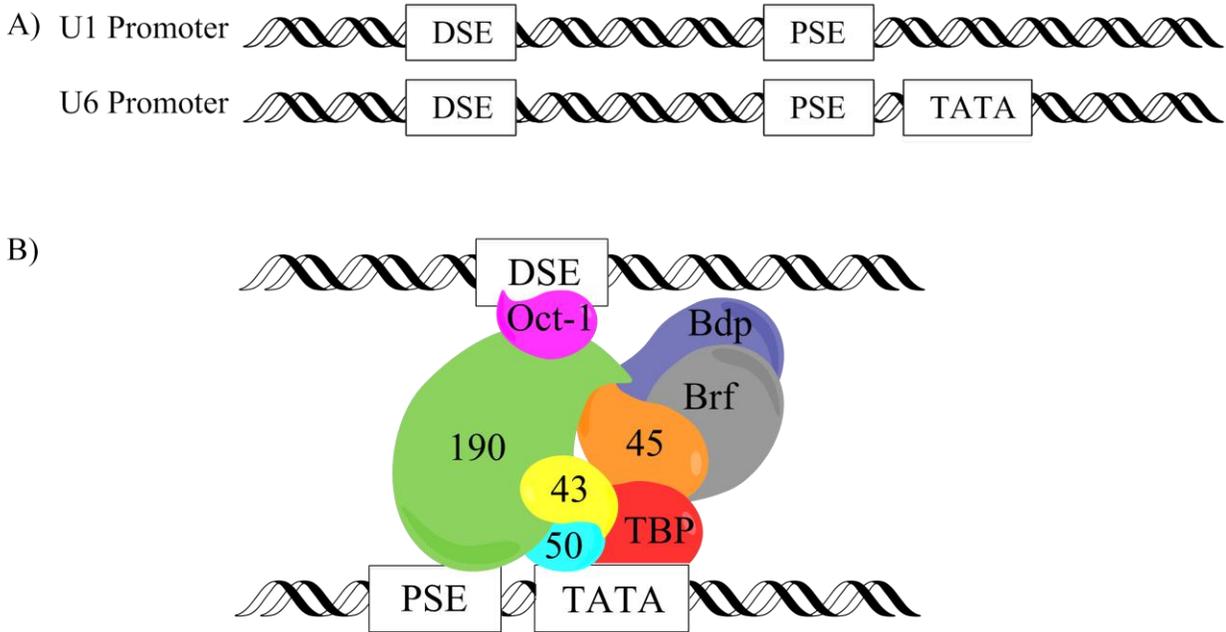


Figure III.1 A) Schematic of the U1 and U6 promoters involved in RNA Polymerase II and III transcription initiation, respectively, adapted from Hernandez et. al. (7). B) Representation of the PIC of SNAPc in RNAP Polymerase III recruitment to the U6 promoter adapted from Hanzlowsky et. al. (14).

This shortened construct could still bind SNAP50, SNAP43 and SNAP19. It was also able to bind to DNA & recruit TBP, Brf2 and Bdp1 to the promoter (23). Attempts to crystallize the individual subunits failed. However, this mini-SNAPc construct offered promise as the subunits appeared to be more stable as a complex than individually. To produce enough protein for crystallographic studies, these four subunits of SNAPc were co-expressed in *E. coli*. Three different plasmids were created containing the open reading frames of the proteins. Sequential transformation of the plasmids into competent cells allowed for protein complex production. The four subunits were then purified from whole cell extracts via the GST tag on SNAP190. This multi-subunit complex was coined “Mini Small Nuclear RNA Activating Protein Complex” (mSNAPc) (24).

III.2. Experimental Procedures

III.2.1. Co-expression of mSNAPc

Three plasmids containing the four open reading frames (ORF) of SNAP190 (1-505) (pGST), SNAP50/SNAP43 (pCDF) and SNAP19 (pRSF) (24) were sequentially transformed into B121 CodonPlus (Stratagene) *E. coli* cells to produce mini-SNAPc (mSNAPc). Initially the SNAP190 plasmid bearing an N-terminal GST affinity tag with a thrombin linker expressing amino acids 1-505 was transformed. The cells were then streaked on ampicillin and chloramphenicol containing agar plates. Colonies containing the ampicillin resistance were selected and made competent using calcium chloride treatment. The SNAP50 and SNAP43 containing plasmid was then transformed into these competent cells. After streaking on ampicillin, chloramphenicol and streptomycin containing agar plates, colonies were selected. These cells were again made competent via calcium chloride treatment and the SNAP19 containing plasmid was transformed. After streaking on ampicillin, chloramphenicol, streptomycin and kanamycin containing agar plates the resulting colonies were selected, grow in 5 mL of Luria Broth (LB) supplemented with 50 mg/mL ampicillin, 50 mg/mL chloramphenicol, 20 mg/mL streptomycin and 20 mg/mL kanamycin at 37°C over night. To the broth was added enough glycerol to bring the concentration up to 20% before aliquots of 1 mL were frozen in liquid nitrogen and stored at -80°C.

For each liter of cells grown, one 250mL flask containing 50mL of LB supplemented with 50mg/mL ampicillin, 50mg/mL chloramphenicol, 20mg/mL streptomycin and 20mg/mL kanamycin was inoculated with 150µL of the glycerol stock and allowed to grow overnight at 37°C. Each 250mL flask was then added to 1L of LB supplemented with 50mg/mL ampicillin, 50mg/mL chloramphenicol, 20mg/mL streptomycin and 20mg/mL kanamycin and allowed to

grow at 37°C until the optical density (OD) reached 0.8 – 1.0. The broth was then cooled to 16°C and protein production was induced by adding 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). The flasks were then left over night before being harvested via centrifugation at 5,000 RPM for 20 minutes. The cell pellets were collected and stored at -20 °C until use.

III.2.2. Initial Standard Purification of mSNAPc

Each liter of harvested cells were suspended in 80mL of HEMGT Buffer (20 mM HEPES, pH 7.9; 2 mM EDTA, 20 mM magnesium chloride, 10% glycerol, 1% Tween 20, 3 mM DDT and 250 mM KCl) with the typical batch being 6 liters of cells total. One Complete Mini Protease Inhibitor Tablet (Invitrogen) was added per 50mL of buffer. Cells were lysed on ice using a Branson Sonifier (3/8 inch tip) with 3 cycles of 1 minute on/1 minute off. The crude cell extract was then clarified at 7,000 RPM for 1 hour. The resulting supernatant was allowed to bind to 10mL Glutathione Sepharose 4B Resin (GST Resin, GE Healthcare) overnight. The resin was then collected by centrifugation at 3,000 RPM and washed with the same buffer.

Collection of protein was achieved by adding 20 units of bovine thrombin per 5mL of resin with cutting occurring over night at 4°C, or by eluting the resin with 50 mM reduced glutathione containing HEMGT Buffer followed by digestion with 20 units of thrombin per 10 mL of elution buffer overnight at 4°C. Purity of the crude fractions was determined by SDS Page.

The crude fractions were then concentrated using 70% ammonium sulfate to 1-2 mL total volume. Further purification was achieved using a Sephadex 200 gel filtration column (1 mL/min flow rate, ~25 mL bed volume, GE Healthcare).

III.2.3. Alternative SNAPc Purification Protocol I

The second purification protocol for mSNAPc was as follows: 3L worth of cells were suspended in 80mL of TEG-250 (20 mM Tris, 2 mM EDTA, 10% glycerol, 3 mM DTT, 250 mM KCl) supplemented with 1 Complete protease inhibitor tablet. The cells were then lysed with a sonicator and the mixture clarified via centrifugation for 40 minutes at 7,000 RPM. 10 mL of GST-resin was added per 40 mL worth of cells and this suspension was allowed to shake over night.

The following day, the resin was collected via centrifugation and washed 3 times with TEG-250 until no protein was detected in the wash. To a second 50 mL of TEG-250 was added reduced glutathione to a concentration of 50 mM. This was used in 10 mL fractions to elute the protein from the resin until no detectable protein was eluting as seen using the Bradford method. Crude protein was collected and frozen with liquid nitrogen for -80°C for storage.

III.2.4. GST Resin Binding Time Optimization

Growth of cells and production of crude lysate were followed as described in Sections III.2.1 and III.2.2. To the 80mL of crude lysate was added 10mL of GST resin. Samples of the resin amounting to 50uL each were removed at 1 hour time points for 4 hours followed by a collection after the resin had been in contact with the lysate overnight. The samples were then washed with the appropriate buffer and resolved on a SDS-Page gel.

III.2.5. Optimization of Length of Time of Thrombin Digestion

Growth of cells and production of crude mSNAPc was followed as described in sections III.2.1 and III.2.2 One milliliter samples of freshly collected mSNAPc were digested with 2 units

of thrombin each with no magnesium chloride in the buffer. Samples were removed and immediately prepped for analysis by SDS-Page at 1 hour time points for 8 hours.

III.2.6. Further Optimization of Buffers

Various buffers were tested for their abilities to suspend the precipitated protein. Purification of the protein was carried out as described above. The buffer compositions as well as the concentrations of the resulting suspensions are reported in **Table III.2.6.1** below.

Table III.1 Buffers used in the optimization of the purification of mSNAPc.

	TEG-250	TTEG-250	TEGG-250	TEG-500	TG-250
Tris	20 mM	40 mM	20 mM	20 mM	20 mM
EDTA	2 mM	2 mM	2 mM	2 mM	
Glycerol	10%	10%	20%	10%	10%
DTT	3 mM				
KCl	250 mM	250 mM	250 mM	500 mM	250 mM
pH	7.9	7.9	7.9	7.9	7.9
Initial Concentration	0.181 mg	0.180 mg	0.165 mg	0.198 mg	0.177 mg
Concentration after centrifugation at 12K RPM	0.148 mg	0.099 mg	0.137 mg	0.157 mg	0.132 mg
Percent of Protein Lost during centrifugation	18%	44%	17%	21%	25%
Concentration after suspension	0.018 mg	0.005 mg	0.006 mg	0.000 mg	0.014 mg

III.2.7 Alternative SNAPc Purification Protocol II

Each liter of harvested cells were suspended in 50mL of HEMGT Buffer (20 mM HEPES, 2 mM EDTA, 20 mM MgCl₂, 10% glycerol, 1% Tween 20, 3 mM DTT, 250 mM KCl, pH 7.9) with the typical batch being 6 liters of cells total. One Complete Protease Inhibitor Tablet (Invitrogen) was added per 50 mL of buffer. Cells were lysed on ice using a Branson Sonifier with 3 cycles of 45 second on/45 seconds off. The crude cell extract was then clarified at 6,000 RPM for 20 minutes. The resulting supernatant was allowed to bind to 5 mL GST Resin per 50 mL of clarified lysate for approximately 4 hours. The resin was then collected by centrifugation at 3,000 RPM and washed with the same buffer.

Collection of crude protein was achieved by adding 20 units of bovine thrombin per 5 mL of resin, or by eluting the resin with 10mM Glutathione containing buffer. Purity of the crude fractions was determined by SDS Page. Further purification was achieved using a Sephadex 200 gel filtration column (1 mL/min).

III.2.8. Crystallization Trials

The optimized purification was performed 25 times with a total of 150 liters of cell culture. The final yield of purified mSNAPc γ 4 was 800 μ L at 3.1 mg/mL. An additional 800 μ L at 8.1 mg/mL was recovered from the void peak of the gel filtration.

Complete crystallization screens of purified mSNAPc revealed several conditions that produced needle-like clusters, the best of which was 8-12% PEG 5000 MME, 100 mM Magnesium Chloride, 10-100 mM Tris pH 8.5, and 100 mM Sodium Chloride.

III.2.9. His-tagged SNAP50

Mutagenesis was employed to introduce six non-cleavable histidines to the N- and C-termini of the SNAP50 subunit, respectively. The histidines were added in two steps with three being added at a time. Insertion was accomplished using a standard Polymerase Chain Reaction (PCR) procedure as described here. For each individual mutation four different combinations of DNA: Primer were sampled, with the ratios being 1 μ L: 100 ng; 2 μ L: 100 ng; 1 μ L: 200 ng; 2 μ L: 200 ng. To each sample was then added 5 μ L of 10xPFU Buffer (Stratagene, La Jolla, CA), 1.5 μ L of 50 mM magnesium chloride (Invitrogen, Carlsbad, CA), 200 μ M of dNTPs (Promega, Madison, WI), 39 μ L of water and 1.25 U of PFU Turbo (Stratagene, La Jolla, CA). The samples were then run through the following sequence provided by Craig Hinkley using a GeneAmp PCR 2400 System (Perkin Elmer). The first cycle was 95°C for 5 minutes. The second cycle was repeated 18 times with the first step being 95°C for 2 minutes, 49°C for 2 minutes and 72°C for 10 minutes. The final cycle was 72°C for 10 minutes.

To introduce the C-terminal tag, primer (5'-GTT GAT CCT GGA ACC TTT AAT CAT CAT CAT TAA GAG CTC GGC GCG CCT-3') (JHG301) along with its reverse complement (JHG302) was subjected to PCR in the presence of the original pCDF vector containing both the SNAP50 and SNAP43 subunits. Insertion was confirmed by sequencing (JHG587). The second set of histidines were added as above using the primer (5'-GGA ACC TTT AAT CAT CAT CAT CAC CAC CAC TAA GAG CTC GGC GCG CCT G-3') (JHG307) and its reverse complement along with the partially mutated pET vector. Insertion was confirmed by sequencing (JHG606).

To introduce the N-terminal tag, primer (5'-ACT TTA ATA AGG AGA TAT ACC ATG CAT CAT CAT GCT GAA GGA AGC CGA-3') (JHG303) along with its reverse complement

(JHG304) was subjected to PCR in the presence of the original pCDF vector containing both the SNAP50 and SNAP43 subunits. Insertion was confirmed by sequencing (JHG588). The second set of histidines were added as above using the primer (5'-ATA AGG AGA TAT ACC ATG CAC CAC CAC CAT CAT CAT GCT GAA GGA AGC -3') (JHG306) and its reverse complement (JHG305) along with the partially mutated pET vector. Insertion was confirmed by sequencing (JHG610).

III.2.10. Purification of Co-expressed His-tagged SNAP50, GST-tagged SNAP190 (1-505), SNAP43 and SNAP19 (N-His γ 4)

The SNAP19 expressing pRSF plasmid was transformed into BL-21 competent cells already harboring the N-terminal His-tag SNAP50 and untagged SNAP43 expressing pCDF plasmid and GST tagged SNAP190 (1-505) expressing pGST plasmid. For a typical purification of culture prepared as described above using HEG 1K buffer 1.5mg/mL of crude SNAP50/SNAP190(1-505)/SNAP43/SNAP19 (N-His γ 4) complex was collected via thrombin digestion off of GST resin.

Additionally Nickel NTA Resin was utilized to purify the crude SNAP50/SNAP190 (1-505)/SNAP43/SNAP19 complex. The protocol was adjusted as follows based on the methods presented in the Nickel NTA handbook: 10 mM imidazole containing sodium phosphate buffer was used during lysis of the cells. The crude lysate typically bound via gravity flow through a column to 15 mL of Ni NTA resin per 50mL of crude lysate. For washing of the Ni NTA resin 20 mM imidazole containing phosphate buffer was used followed by 250 mM imidazole containing sodium phosphate buffer to elute the protein.

III.2.11. SNAP190 (1-131), (1-135), (1-255), (1-260), (1-265)

Mutagenesis was employed to introduce two sequential stop codons (**TAA TGA TCA**) into the SNAP190 open reading frame downstream of the Myb domain. The sequence contained an additional three base pairs that constitute a *BclI* restriction site (TAA **TGA TCA**). Insertion was accomplished using a standard Polymerase Chain Reaction (PCR) procedure as described.

To introduce the stop site at position 131, primer (5'-TCC AAA GGC ACC AAG GTG AAA **TAA TGA TCA** GAT GGC AAA AGC CTG CCC CCA-3') (JHG327) along with its reverse complement (JHG328) was subjected to PCR in the presence of the original pGST vector containing the SNAP190 (1-505) open reading frame. Insertion was confirmed by digestion with *BclI* as well as sequencing (JHG680).

To introduce the stop site at position 135, primer (5'-AAG GTG AAA GAT GGC AAA AGC **TAA TGA TCA** CTG CCC CCA AGC ACA TAC ATG-3') (JHG325) along with its reverse complement (JHG326) was subjected to PCR in the presence of the original pGST vector containing the SNAP190 (1-505) open reading frame. Insertion was confirmed by digestion with *BclI* as well as sequencing (JHG642).

To introduce the stop site at position 255, primer (5'-CCA GAA GAG GCC TTG CTG GGA **TAA TGA TCA** AAC AGG CTG GAC AGC CAC GAC-3') (JHG319) along with its reverse complement (JHG320) was subjected to PCR in the presence of the original pGST vector containing the SNAP190 (1-505) open reading frame. Insertion was confirmed by digestion with *BclI* as well as sequencing (JHG641).

To introduce the stop site at position 260, primer (5'-CTG GGA AAC AGG CTG GAC AGC TAA TGA TCA CAC GAC TGG GAG AAG ATT TCC AAT ATT-3') (JHG321) along with its reverse complement (JHG322) was subjected to PCR in the presence of the original

pGST vector containing the SNAP190 (1-505) open reading frame. Insertion was confirmed by digestion with BclI as well as sequencing (JHG672).

To introduce the stop site at position 265, primer (5'-GAC AGC CAC GAC TGG GAG AAG TAA TGA TCA ATT TCC AAT ATT AAC TTT GAA-3') (JHG323) along with its reverse complement (JHG324) was subjected to PCR in the presence of the original pGST vector containing the SNAP190 (1-505) open reading frame. Despite several trials, insertion was unsuccessful.

Expression of these truncated SNAP190 constructs was achieved both individually and as a complex with SNAP50 and SNAP43 via co-expression as described previously. Cell cultures were typically grown in terrific broth to an $OD_{600} = 1.5$ and induced with IPTG at room temperature.

Each 1.5 L of cell culture was suspended in 100 mL of HEGT 250 plus one Complete protease inhibitor tablet. The mixture was sonicated for three cycles of 45 seconds with 45 seconds elapse between each cycle. The sonicated mixture was then clarified at 7,000 RPM for 20 minutes. The decanted crude lysate was then combined with 5mL 50:50 glutathione resin in HEGT 250. The slurry mixed for 3.5 hours at 4°C. Resin was collected by centrifugation at 3,000 RPM for 10 minutes, decanted and washed with 40mL HEGT 250 until a Bradford reading of <0.05 was achieved. Crude protein was digested from the resin by addition of 100U of thrombin to each sample at 4°C for one hour. Resin was again pelleted as described and the protein containing supernatant collected. Resin was washed with 5mL portions of HEGT 250 until a Bradford reading of <0.10 was observed. Crude samples were stored at -80°C by flash freezing with liquid nitrogen.

III.2.12. Tagless SNAP190 (1-505)

Mutagenesis was employed to introduce a NdeI restriction site to the N-terminus of the SNAP190 ORF using primer (5'-CTG GTT CCG CGT GGC TCT CAT ATG GAT GTA GATGCT GAA-3')(JHG338) and its reverse complement (JHG339) where a XbaI site had existed previously (**Figure III.2**). Insertion was accomplished using a standard PCR procedure.

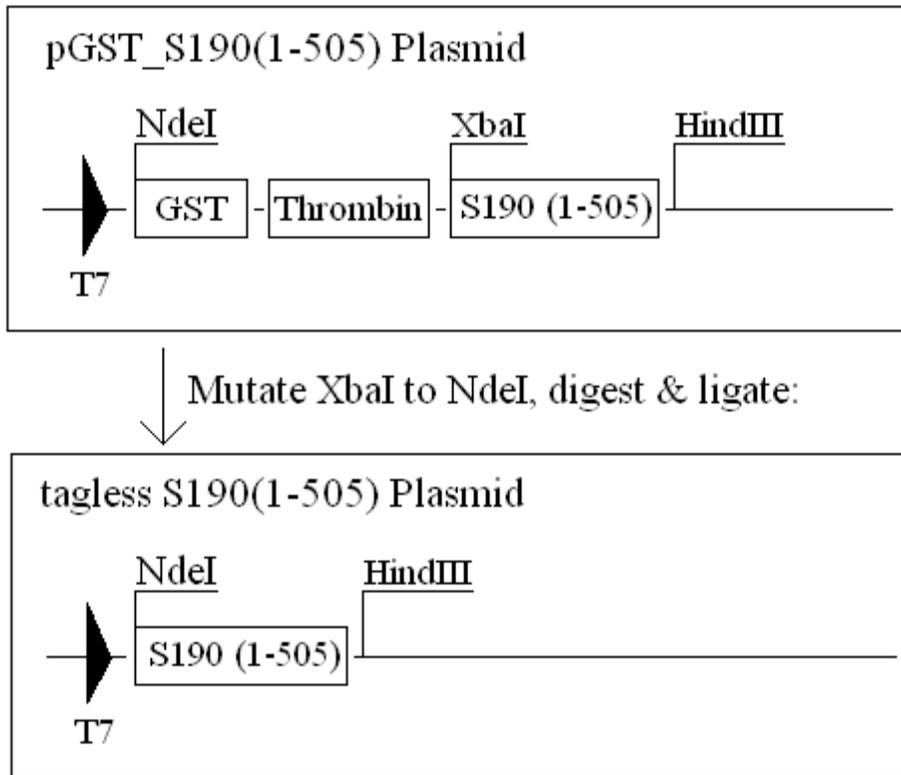


Figure III.2 Schematic of the cloning for the ORF of tagless SNAP190 (1-505)

For each individual mutation four different combinations of DNA: Primer were sampled, with the ratios being 1 μ L:100ng; 2 μ L: 100ng; 1 μ L: 200ng; 2 μ L:200ng. To each sample was then added 7.5 μ L of 10xPFU Buffer (Stratagene, La Jolla, CA), 1.0 μ L of 50 mM magnesium chloride (Invitrogen, Carlsbad, CA), 200 μ M of dNTPs (Promega, Madison, WI), 36.5 μ L of

water and 1.25 U of Pfu Turbo (Stratagene, La Jolla, CA). The samples were then run through the following sequence using a GeneAmp PCR 2400 System (Perkin Elmer). The first cycle was 95°C for 2 minutes. The second cycle was repeated 30 times with the first step being 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 10 minutes. The final cycle was 72°C for 10 minutes.

The crude PCR mixture was digested with *DpnI* for one hour at 37°C before transformation into DH5 α competent cells. DNA was extracted from colonies forming after 18 hours of incubation at 37°C on ampicillin containing agar plates using a QIA Miniprep Kit. Eight μ L of the DNA was digested with 1 μ L of *NdeI* in the presence of 1 μ L of Buffer 2 (New England Biolabs) for one hour before being run on a 1% agarose gel.

The open vector was extracted from the gel using a QIA gel extraction kit. 15 μ L of the vector was ligated together using 4 μ L of 5x T4 ligation buffer and 1 μ L of the T4 DNA ligase for one hour before being transformed into DH5 α competent cells. Colonies forming after 18 hours of incubation at 37°C on ampicillin containing agar plates were grown ON in LB supplemented with 50 μ g/mL ampicillin before the plasmid was prepped out using a QIA Miniprep Kit. The vector's completeness was confirmed through sequencing (JHG699).

II.2.13. Maltose Binding Protein Tagged SNAP190 (1-505)

The empty vector pMAL-c4x (New England Biolabs) contains the maltose binding protein open reading frame with a C-terminal Factor Xa cleavable linker. Since the SNAP190 ORF already contained flanking *XbaI* and *HindIII* restriction sites, it could be directly ligated into the pMAL-c4x vector. First, 24 μ L of the pGST vector containing the SNAP190 (1-505) ORF and 24 μ L of the pMAL-c4x vector were each digested with 1 μ L of BSA, 1 μ L of *XbaI*, 1 μ L of *HindIII*, and 3 μ L of Buffer 2 (New England Biolabs) for 6 hours. The digests were then run

on a 1% agarose gel after which the ORF of SNAP190 and the open vector of pMAL-c4x were recovered using a QIA Gel Extraction Kit.

The 11 μ L of the SNAP190 ORF and 4 μ L of the empty vector were ligated together using 4 μ L of 4x T4 ligase buffer and 1 μ L of T4 DNA ligase for 1 hour. The resulting ligation reaction was transformed into DH5 α competent cells (Invitrogen) and streaked onto ampicillin containing agar plates which were allowed to grow ON at 37°C.

Subsequent colonies were selected and allowed to grow ON in 5mL of LB supplemented with 50 μ g/mL ampicillin. DNA was then extracted using the QIA Miniprep Kit. Sequencing was used to confirm the successful insertion of the reading frame (JHG687).

E. coli cell cultures were grown in LB the presence of 100 μ g/mL ampicillin and 50 μ g/mL chloramphenicol at 37°C until OD₆₀₀= 0.8 to 1.0. Induction with IPTG occurred at 16°C and cell cultures were allowed to grow overnight. Cells were collected by centrifugation at 5,000 RPM for 20 minutes. Cell pellets were stored at -20°C.

Purification began with suspension of 2 L of cells in 50mL TEG 200 (20 mM Tris-HCl, pH 7.4, 0.2 M NaCl, 10 mM β -mercaptoethanol, 1 mM EDTA, and 10% glycerol) supplemented with one Complete protease inhibitor tablet. Cells were lysed by sonication in 30 second cycles with 30 second rests between cycles. The crude lysate was clarified by centrifugation at 7,000 RPM for 20 minutes. The clarified supernatant was decanted over a 1 mL amylose resin column at 4°C. The column was washed with 10 column volumes of TEG 200 and the protein eluted with the above buffer plus 10mM maltose. Crude protein samples were stored at -80°C by freezing with liquid nitrogen.

III.2.14 Maltose Binding Protein Tagged SNAP190 (1-505) with Thrombin

Linker

The Factor Xa protease cleavage site on the maltose binding protein tagged SNAP190 (1-505) was mutated to a Thrombin protease cleavage site using primer (5'- CAA TAA CAA TAA CAA CAA CCT CCT GGT TCC GCG TGG CTC TAG AGA ATT CGG ATC CTC TAG AGT C -3') (JHG335) and its reverse complement (JHG336). Insertion was accomplished using a standard PCR procedure as described here.

Four different combinations of DNA: Primer were sampled, with the ratios being 1 μ L:100ng; 2 μ L:100ng; 1 μ L:200ng; 2 μ L:200ng. To each sample was then added 7.5 μ L of 10xPFU Buffer (Stratagene, La Jolla, CA), 1.5 μ L of 50 mM magnesium chloride (Invitrogen, Carlsbad, CA), 200 μ M of dNTPs (Promega, Madison, WI), 39 μ L of water and 1.25 U of PFU Turbo (Stratagene, La Jolla, CA). The samples were then run through the following sequence using a GeneAmp PCR 2400 System (Perkin Elmer). The first cycle was 95°C for 2 minutes. The second cycle was repeated 30 times with the first step being 95°C for 30 seconds, 45°C for 30 seconds and 72°C for 10 minutes. The final cycle was 72°C for 10 minutes. Mutation was confirmed by non-digestion with XmnI as well as sequencing (JHG689).

Purifications were employed as described above for the factor Xa linker SNAP190 with the addition of 1% Tween 20 to the TEG 200 buffer.

III.2.15. Maltose Binding Protein Tagged SNAP190 (1-505) with Smt3 Linker

A plasmid containing the ORF of the Smt3 protein was generously provided by Christopher Lima. There existed a BamHI site on the C-terminus but no restriction site at the N-terminus (**Figure III.3**). The Factor Xa recognition sequence was preceded by a SacI restriction

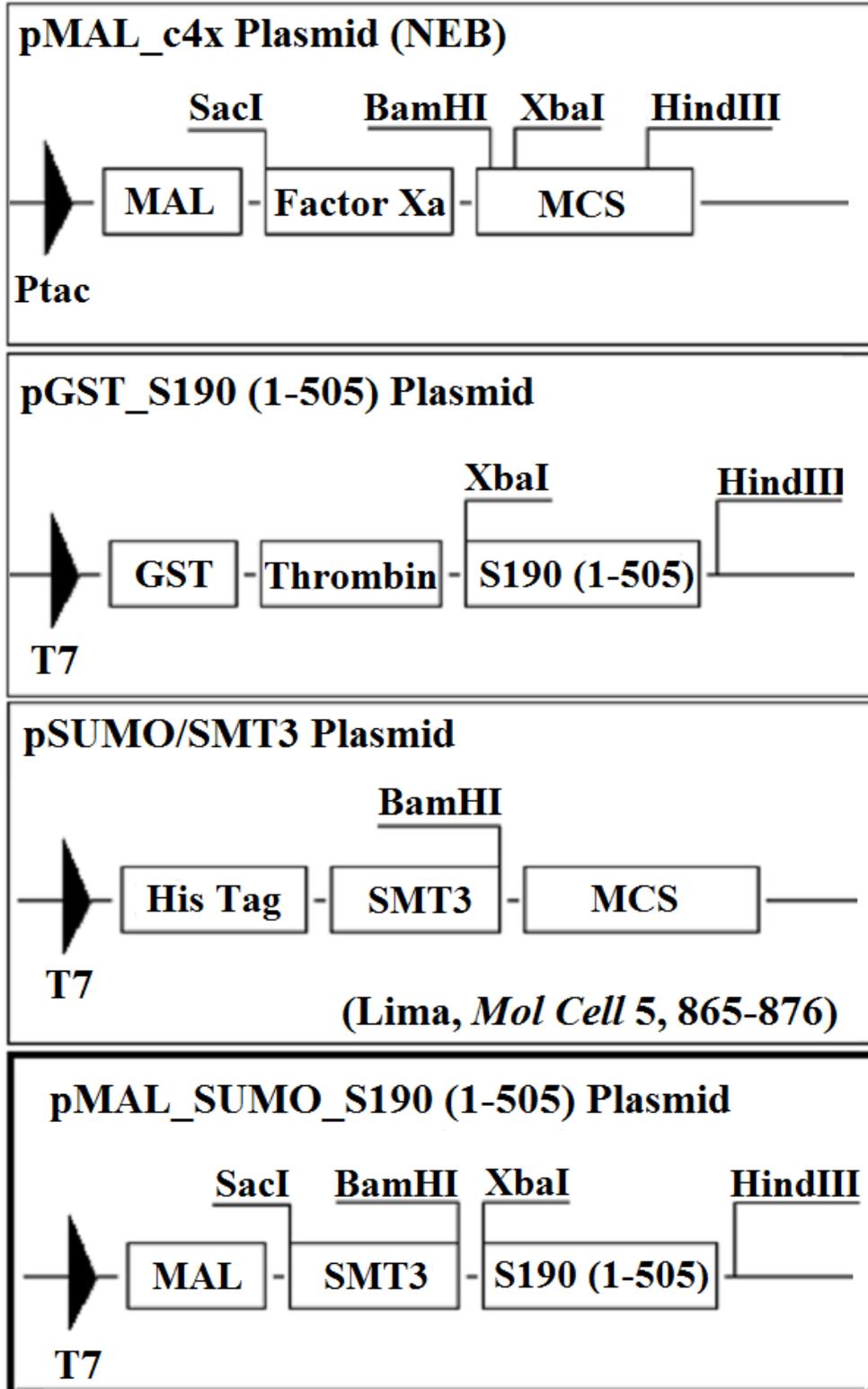


Figure III.3 Schematics of the pMal-c4x, pGST_190 (1-505), pSUMO/SMT3, and pMAL_SUMO_S190 (1-505) showing the relevant restriction sites.

site in the pMAL_c4x plasmid. Therefore a Sac1 restriction site was introduced to the N-terminus of the Smt3 reading frame on the original plasmid provided by Lima. This was accomplished using the primer (5'- AGC AGC CAT CAT CAT CAT GAG CTC AGC AGC GGC CTG GTG C -3') (JHG340) and its reverse complement (JHG341). Mutation was accomplished using PCR as described in section II.1.7 with the exception that the annealing temperature was changed to 63°C in the second step.

The crude PCR reaction was processed as described previously. Twenty four microliters of the prepped DNA was digested with 1µL of Sac1, 1µL of BAMHI, 1µL of BSA, and 1µL of Buffer 1 (New England Biolabs) for 1 hour at 37°C. The digest was run on a 1% DNA agarose gel and the Smt3 ORF was extracted from the gel using the QIA gel extraction kit (Qiagen). The pMAL_c4x plasmid was similarly digested and extracted. The open vector (3.75µL) and the insert (11.25µL) were ligated together using 4µL of 5x T4 ligation buffer and 1µL of T4 DNA ligase. Insertion was confirmed by digestion with Ava1 and sequencing (JHG690).

Into the new vector, known as the pMAL_SUMO vector, was ligated the SNAP190 (1-505) ORF as described previously. This new construct was known as the PS_190 for pMAL_SUMO_SNAP190 (1-505). Improper insertion between the maltose binding protein and the Smt3 protein was confirmed by sequencing (JHG711 & JHG712). To correct the sequence mutagenesis was again employed using primer (5'- GCG CAG ACT AAT TCG GAG CTC AGC AGC GGC CTG-3') (JHG354) and its reverse complement (JHG355) and PCR as described above with the exception that the annealing temperature was changed to 51°C in the second step. The correct sequence was confirmed by sequencing (JHG724). The empty pMAL_Smt3 vector was similarly corrected (JHG762).

III.2.16. Maltose Binding Protein Tagged SNAP190 (1-131) with Smt3 Linker

Into the corrected pMAL_SUMO vector (Section III.2.15) was ligated the ORF of SNAP190 (1-131) the creation of which is described in section II.1.4. The ligation was achieved following the exact protocol laid out in section II.1.8 for the insertion of SNAP190 (1-505).

III.2.17 Maltose Binding Protein Tagged SNAP50 with Smt3 Linker

The fusion of maltose binding protein to the Smt3 protein was added to the N-terminus of SNAP50 on the pCDF vector containing originally the untagged SNAP50 and SNAP43 ORF (**Figure III.4**). First, the pMal_SUMO empty vector was mutated just past the C-terminus of the Smt3 protein to introduce a Nco1 restriction site using primer (5' - CAC AGA GAA CAG ATT GGT ACC ATG GCT AGA GTC GAC CTG CAG -3') (JHG374) and its reverse complement (JHG 375) and PCR as described in section 1.2.1.17 with the exception that the annealing temperature was changed to 67°C in the second step. Mutation was confirmed by successful digestion of the plasmid with Nco1 as well as sequencing (JHG823).

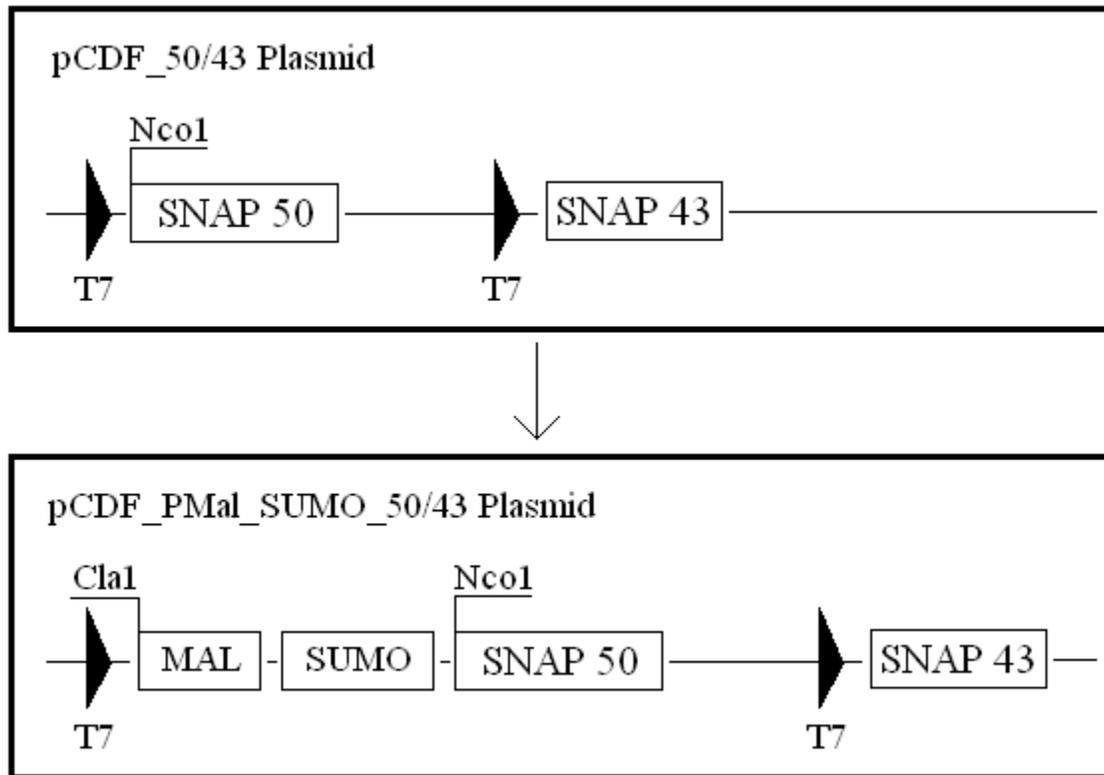


Figure III.4 Schematic of the pCDF_50/43 and pCDF_pMAL_SUMO_50/43 plasmids showing relevant restriction sites.

Next, the Nco1 mutated pMal_SUMO vector was further mutated just upstream of the N-terminus of the maltose binding protein ORF to introduce a Cla1 restriction site using primer (5'-CAC CAA CAA GGA CCA TAA TCG ATG AAA ATC GAA GAA GGT AAA -3') (JHG372) and its reverse complement (JHG373). PCR was used as described in section above with the exception that the annealing temperature was changed to 65°C in the second step. Mutation was confirmed by successful digestion of the plasmid with Cla1 followed by sequencing (JHG822).

Flanking Cla1 and Nco1 sites were introduced to the N-terminus of the SNAP50 ORF in the pCDF plasmid via mutagenesis using primer (5'- TAA CTT TAA TAA GGA GAT AAT CGA TGG GGA CCA TGG CTG AAG GAA GCC GA -3') (JHG391) and its reverse complement (JHG 392). PCR was used as described above with the exception that the annealing

temperature was changed to 65°C in the second step and a MJ Research PTC-100 thermal cycler was used. Successful mutation was confirmed by digestion with Cla1.

The Cla1/Nco1 mutated pCDF vector (24µL) as well as the Cla1/Nco1 mutated pMal_SUMO vector (24µL) were double digested independently with 1µL of Cla1, 1µL of Nco1, 1µL of BSA and 3µL of Buffer 4 (New England Biolabs) for 1 hour at 37°C. The digests were run on a 1% agarose DNA gel and the appropriate bands excised. DNA was extracted from these bands using a QIA gel extraction kit. The open vector (3.75µL) and the insert (11.25µL) were then ligated together with 1µL of T4 DNA ligase and 4µL of the corresponding 5x T4 DNA ligase buffer for one hour at room temperature. Insertion was confirmed by sequencing (JHG860, JHG861, and JHG862).

Sequencing revealed an error in the pCDF promoter region as well as the ORF of the SUMO protease cut site. The first error was corrected using mutagenesis via primer (5' - ACC TTC TTC GAT TTT CAT GGT ATA TCT CCT TAT TAA AGT TA -3') (JHG407) and its reverse complement (JHG408). The correction was confirmed by sequencing (JHG920). The SUMO protease cut site was corrected using a similar method initially using primer (5' - CAC AGA GAA CAG ATT GGT GGA ATG GCT GAA GGA AGC CG -3') (JHG418) and its reverse complement (JHG419) with the only change to the protocol as described previously being an elongation temperature of 68°C. The partial correction was confirmed by sequencing (JHG946). The complete correction was achieved via mutagenesis as described previously in section III.2.14 via primer (5' - GCT CAC AGA GAA CAG ATT GGT GGA AGT ATG GCT GAA GGA AGC CGA GG -3') (JHG416) with the correction being confirmed by sequencing (JHG954).

As the Maltose Binding Protein Tagged SNAP50 with Smt3 Linker plasmid did not produce any of the desired protein, the N-terminal region was again mutated based on the known sequence of the pCDF plasmid. Mutagenesis was used as described previously with primer (5' - AAC TTT AAT AAG GAG ATA TAC CAT GAA AAT CGA AGA AGG T -3') (JHG698) and its reverse complement (JHG699). The correction to the N-terminus of the ORF was confirmed by sequencing (JHG1574).

III.2.18. SNAP19 Truncations

First, the SNAP19 ORF had to be corrected due to a point mutation (V2L) using primer (5' - ATA AGG AGA TAT ACC ATG CTG AGC CGG CTT CAG GAA C -3') (JHG397) and its reverse complement (JHG 398). Insertion was accomplished using a standard PCR procedure as described with the exception that the annealing temperature was changed to 55°C and the elongation time shortened to 7 minutes in the second step and a MJ Research PTC-100 thermal cycler was used. Successful mutation was confirmed by sequencing (JHG 873).

Stop codons were then inserted into the SNAP19 ORF to produce two truncations: SNAP19 (1-41) and SNAP19 (1-85). This was achieved using PCR as described above and primer (5' - CTC CAA TCA ATG ATC AGT TAA TAG TCT AGA AGA GGG GAT -3') (JHG393) and its reverse complement (JHG 394) as well as primer (5' ACA AAG AGT CAT GTG ACG TAA TAG GAA GAG GAG GAG GAG GAA -3') (JHG395) and its reverse complement (JHG 396) respectively. Sequencing (JHG883 and JHG892) confirmed the mutations.

III.2.19 Surface Entropy Reduction Mutations of SNAP190 (1-505)

Surface Entropy Reduction calculations were performed on the SNAP190 (1-505) sequence using the Surface Entropy Reduction Prediction Server at the Eisenberg Laboratory, UCLA (<http://nihserver.mbi.ucla.edu/SER/>), revealing several a possible sites of mutation. The selected mutant of K294A, Q295A and E296A was achieved via mutagenesis using the above described protocol with primer (5'- TCG GAG CAC CCC AGC ATC AAC GCG GCA GCT TGG AGC AGG GAG GAG GAG -3') (JHG414) and its reverse complement (JHG415). The mutation into the pMAL_SUMO_190 plasmid was confirmed by sequencing (JHG931).

III.2.20 Maltose Binding Protein tagged SNAP190 (Δ 131-260) with SMT3 cut site and various linkers

To achieve a SNAP190 ORF expressing amino acids 1-131 and 260-505 as a fusion with various lengths of poly-glycine/serine linking the two, the amino acids between 130 and 260 were removed using a Cla1 restriction site mutation. First the Cla1 restriction site was introduced downstream of amino acid 130 using mutagenesis via primer (5'-TCC AAA GGC ACC AAG GTG AAA ATC GAT GGC AAA AGC CTG CCC CCA -3') (JHG557) and its reverse complement (JHG 558).

Four different combinations of DNA: Primer were sampled, with the ratios being 1 μ L: 100 ng; 2 μ L: 100 ng; 1 μ L: 200 ng; 2 μ L: 200 ng. To each sample was then added 7.5 μ L of 10x PFU Buffer (Stratagene, La Jolla, CA), 1.0 μ L of 50 mM magnesium chloride (Invitrogen, Carlsbad, CA), 200 μ M of dNTPs (Promega, Madison, WI), 38.5 μ L of water and 1.25 U of PFU Turbo (Stratagene, La Jolla, CA). The samples were then run through the following sequence using a PTC-100 Thermal cycler (MJ Research). The first cycle was 95°C for 2 minutes. The

second cycle was repeated 30 times with the first step being 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 14 minutes. The final cycle was 72°C for 10 minutes. Mutation was confirmed by digestion with Cla1.

The second Cla1 restriction site was introduced upstream of amino acid 259 using mutagenesis as described above via primer (5'-TTG CTG GGA AAC AGG CTG GAC **ATC** **GAT** AGC CAC GAC TGG GAG AAG ATT -3') (JHG559) and its reverse complement (JHG 560). Mutation was confirmed by digestion of 24µL of the double mutant plasmid using 1 µL of Cla1 restriction enzyme (NEB), 1µL of BSA, and 3µL of Buffer 4 (New England Biolabs) for 1 hour at 37°C. The digested plasmid was then run on a 1% Agarose gel after which the digested plasmid was recovered using a QIA Gel Extraction Kit.

The 15µL of the digested SNAP190 double mutant plasmid was ligated shut using 4µL of 4x T4 ligase buffer and 1 µL of T4 DNA ligase for 1 hour. The resulting ligation reaction was transformed into DH5α competent cells (Invitrogen) and streaked onto ampicillin containing agar plates which grew colonies ON at 37°C. Colonies were prepped for DNA, and the removal of the desired amino acids confirmed by sequencing (JHG1260).

Removal of the remaining Cla1 restriction site between amino acids 130 and 260 was achieved by mutagenesis using the above protocol and primer (5'- TCC AAA GGC ACC AAG GTG **AAA AGC** CAC GAC TGG GAG AAG ATT -3') (JHG639) and its reverse complement (JHG 640). A test digest with Cla1 along with sequencing confirmed the correct mutation (JHG 1325).

Insertion of a thrombin recognition sequence between amino acids 130 and 260 was achieved by mutagenesis of the above plasmid using the above protocol with the exception that a Techie Thermal cycler was used with a final cycle time of 14 minutes along with primer (5'-

TCC AAA GGC ACC AAG GTG AAA CTG GTT CCG CGT GGC TCT AGC CAC GAC
TGG GAG AAG ATT -3') (JHG641) and its reverse complement (JHG 642). A partially correct
sequence was confirmed by sequencing (JHG 1355). This plasmid was corrected by repeating the
above method on the partially correct plasmid resulting in the final plasmid (JHG1371).

Linkers of the format "GGSGG" were also inserted both upstream and downstream of the
thrombin recognition sequence using the same mutagenesis protocol outlined above. For the
upstream insertion, primer (5' - CTG GTT CCG CGT GGC TCT GGT GGC AGT GGT GGC
AGC CAC GAC TGG GAG AAG ATT TCC -3') (JHG662) and its reverse complement (JHG
663) were used. For the downstream insertion, primer (5' - GGT CCA AAG GCA CCA AGG
TGA AAG GTG GCA GTG GTG GCC TGG TTC CGC GTG GCT CTA GCC AC -3')
(JHG664) and its reverse complement (JHG 665) were used. Sequencing confirmed the
insertions (JHG1427 and JHG1443 respectively).

To obtain the double insertion, the plasmid containing the upstream insertion was
mutated using the primers for the downstream insertion as described above. The double insertion
was confirmed through sequencing (JHG1493).

An additional "GGSGG" sequence was added to the above plasmid using mutagenesis as
described above via primer (5' -TCT GGT GGC AGT GGT GGC GGT GGC AGT GGT GGC
AGC CAC GAC TGG GAG AAG ATT -3') (JHG692) and its reverse complement (JHG693).
The resulting insertion of "GGSGGLVPRGSGSGGGSGG" was confirmed by sequencing
(JHG1565). The resulting plasmid was given to Dr. Stacy Hovde for completion.

III.2.21 Maltose Binding Protein tagged SNAP190 (260-505) with SMT3 cut site

Into the double inserted Maltose Binding Protein tagged SNAP190 (Δ 131-260) with SMT3 cut site was mutated an Xba1 restriction site via mutagenesis as described above. The primer (5'-GGC TCT GGT GGC AGT GGT GGC **TCT AGA** AGC CAC GAC TGG GAG AAG ATT -3') (JHG690) and its reverse complement (JHG691) were used and the correct mutation confirmed by sequencing (JHG 1560). The resulting plasmid was give to Dr. Stacy Hovde where it was digested with Xba1, and the resulting plasmid was then ligated back together.

III.3. Results and Discussion

III.3.1. Purification and attempted crystallization of the SNAPc complex

Initially, four subunits of mSNAPc (SNAP190 1-505, SNAP50, SNAP45 and SNAP19) were co-expressed and co-purified via a Glutathione S-Transferase (GST) affinity tag attached to the N-terminus of SNAP190 via a thrombin recognition sequence. This produced crude protein on a scale of approximately 1mg/liter of cell culture (**Figure III.5**). The purity of the protein was approximately 80% as evident by SDS-Page. However, the stoichiometry of the subunits was in question as the intensity of the bands varied.

To ensure and test the stoichiometry the protein was further purified by gel filtration chromatography. A large portion of the protein was present in the void fraction, suggesting aggregation of the various subunits together (fractions 8-10 in **Figure III.6**).

To reduce the aggregation or increase the desired stoichiometry several buffers were tested that contained various amounts of detergents and salts believed to help reduce unwanted protein-protein interactions that might result in aggregation. Due to the increase in protein collected, the buffer TEG-250 (**Table III.2**) was used for subsequent purifications.

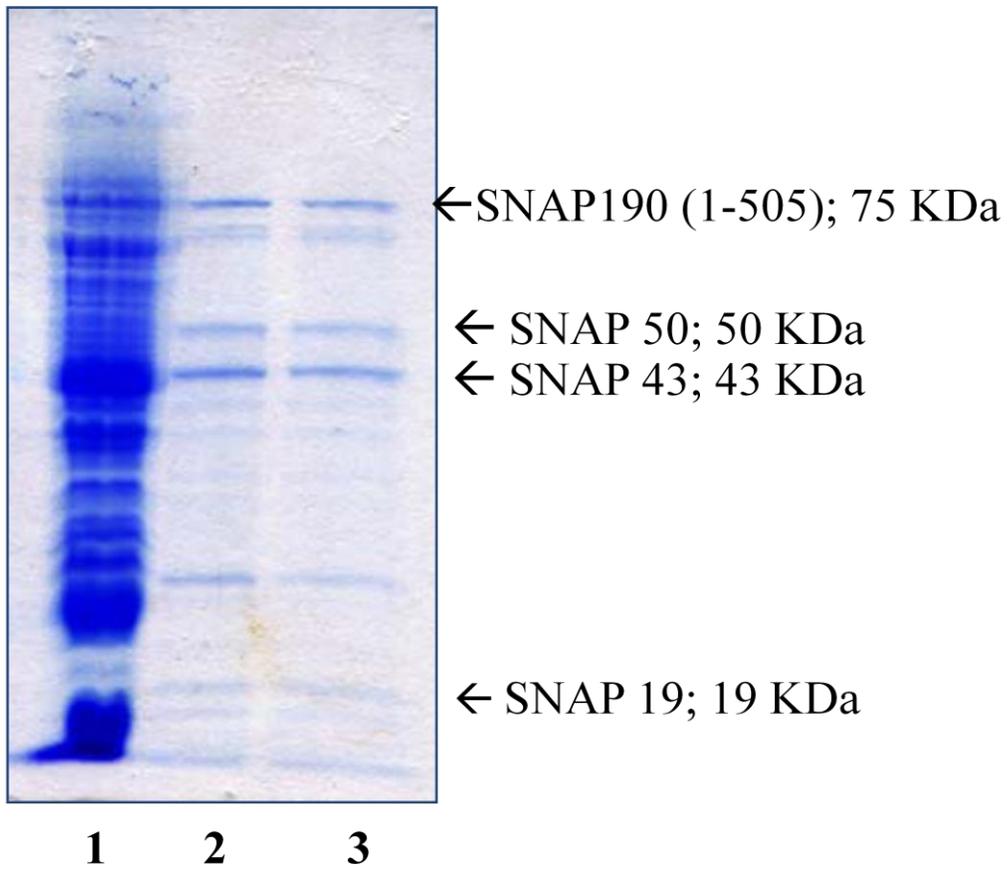


Figure III.5 SDS Page gel. Lane 1: Crude lysate of mSNAPc. Lanes 2, 3: Crude protein elutions of mSNAPc.

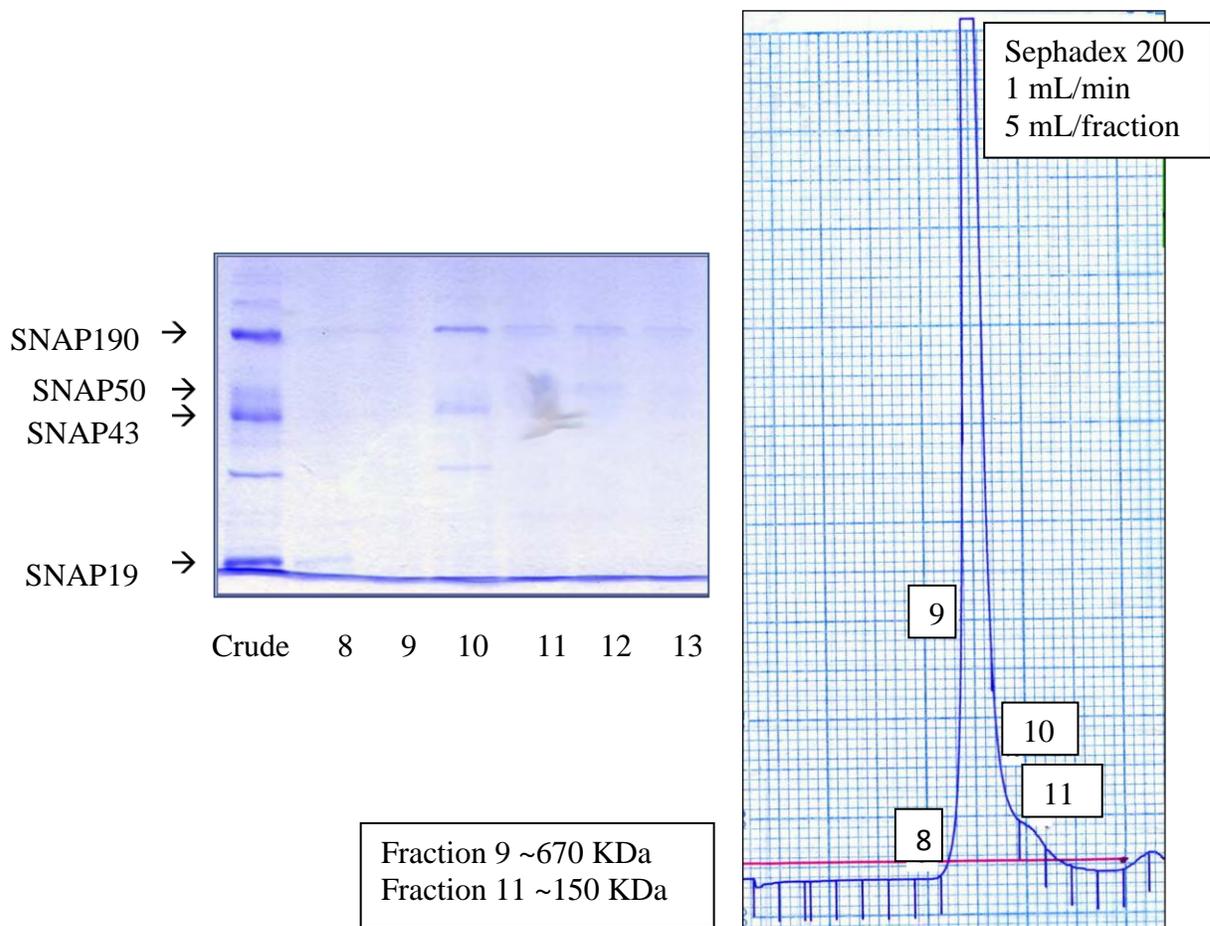


Figure III.6 SDS Page gel and gel filtration chromatograph of crude mSNAPc. Lane Crude: Load of mSNAPc after Ni affinity purification. Lanes 8-12: Fractions 8-12 of the chromatograph for the gel filtration shown to the left.

Table III.2 Buffers used in the purification of mSNAPc.

	TEMG-250	TEG-250	TEMGT-250B	TEMGT-500	TEMGT-1K	TEMGT-pH6.4
Tris	20 mM	20 mM	20 mM	20 mM	20 mM	20 mM
EDTA	2 mM	2 mM	2 mM	2 mM	2 mM	2 mM
MgCl ₂	20 mM	-	20 mM	20 mM	20 mM	20 mM
Glycerol	10%	10%	10%	10%	10%	10%
Tween 20	-	-	1%	1%	1%	1%
DTT	3 mM	3 mM	3 mM	3 mM	3 mM	3 mM
KCl	250 mM	250 mM	250 mM	500 mM	1M	250 mM
pH	7.9	7.9	7.9	7.9	7.9	6.4
Protein collected	0.27 mg	0.49 mg	0.28 mg	0.33 mg	0.25 mg	0.10 mg

Though yield increased, the amount of protein present in the void fraction remained unaffected. Steps were then taken to reduce the amount of time of the purification. In total, from the time the cell cultures were lysed to the time of collection of crude product from the affinity column was a total of three days. During this time, it was possible irreversible aggregation was occurring.

The first step to be optimized was the binding of the protein to the GST resin. GST resin was allowed to interact with the soluble fraction of the cell lysate for various amounts of time, and the total amount of bound protein was determined by SDS-Page analysis. It was found that it took approximately three to four hours for the matrix to become saturated (**Figure III.7**). This reduced the binding time from overnight to four hours.

Next the thrombin digestion was optimized. Thrombin was added to GST affinity purified mSNAPc which had been eluted from the resin, thus it still possessed the N-terminal GST fusion to the SNAP190 subunit. Samples were removed at various time points and the degree of digestion was analyzed by SDS-Page (**Figure III.8**).

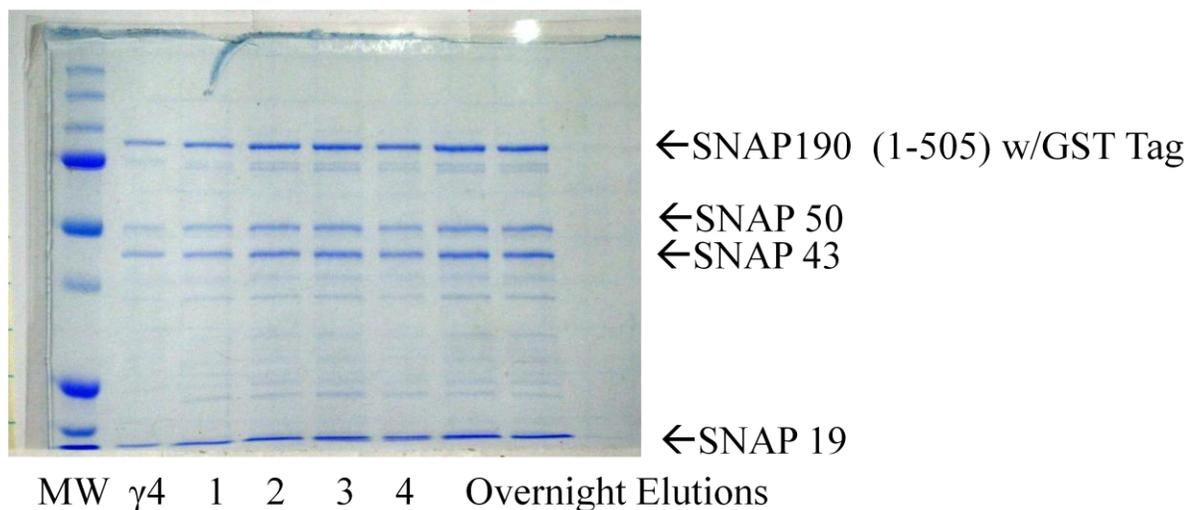


Figure III.7 GST binding optimization of mSNAPc. Lane 1: Molecular weight standard. Lane 2: Crude mSNAPc $\gamma 4$. Lane 3: One hour binding. Lane 4: Two hours binding. Lane 5: Three hours binding. Lane 6: Four hours binding. Lanes 7-8: Overnight binding.

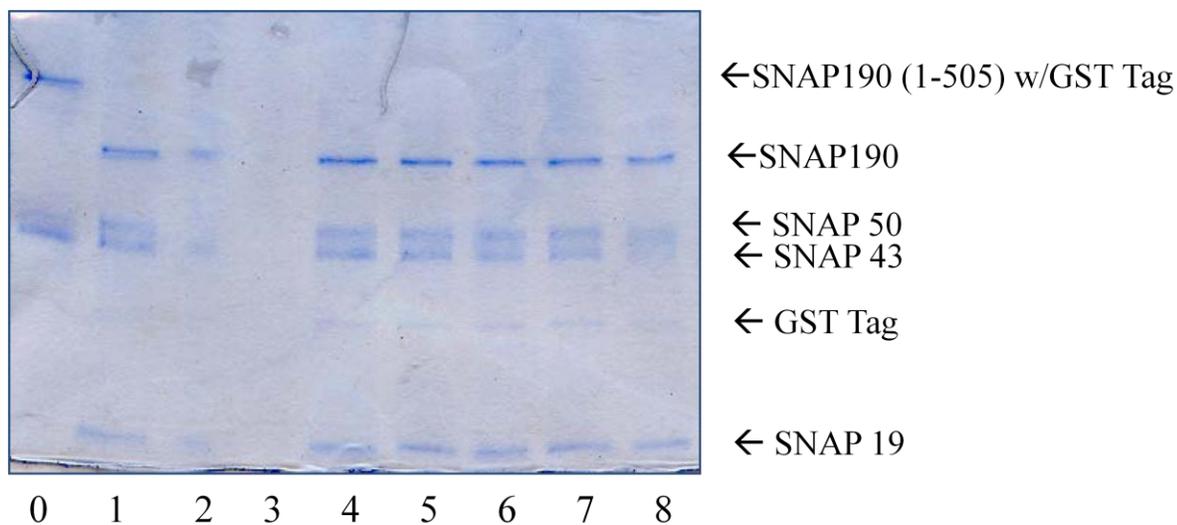


Figure III.8 Gel of thrombin digestion optimization. Lane 0: Crude mSNAPc. Lanes 1-8: One hour increments of digestion with thrombin showing cleavage of the affinity tag.

The gel revealed complete cleavage of the GST tag after less than one hour of digestion. Therefore the protocol was changed to include a 1 hour digestion time, rather than the digestion being completed overnight.

Due to the persistence of aggregation and the difficulty in suspending the ammonium sulfate precipitated protein, buffers were again optimized. It was first determined that 40% ammonium sulfate was sufficient to completely precipitate all protein. However, the precipitated protein was still unable to be dissolved.

Though the concentration of the buffer, EDTA, glycerol and potassium chloride were unable to affect the suspension of the precipitated protein, lowering the concentration of the glutathione elution buffer's concentration from 50mM glutathione to 10mM glutathione corrected the problem.

It was also found that the addition of 0.1% Tween 20 to the buffers allowed for easier concentration using Centricon (Millipore) concentrators with less aggregation observed during gel filtration. Lauryl Maltoside was also tested, but proved to be slightly inferior in preventing aggregation loss of protein during concentration (**Figure III.9**). Addition of spermine or a DNase to the crude fractions further reduced the apparent aggregation of the samples (**Figure III.10**).

These observations led to an optimized purification protocol which produced less aggregated protein (**Figure III.11**). Crystallization trials ensued, but the quality of the crystals were insufficient for x-ray data collection (**Figure III.12**).

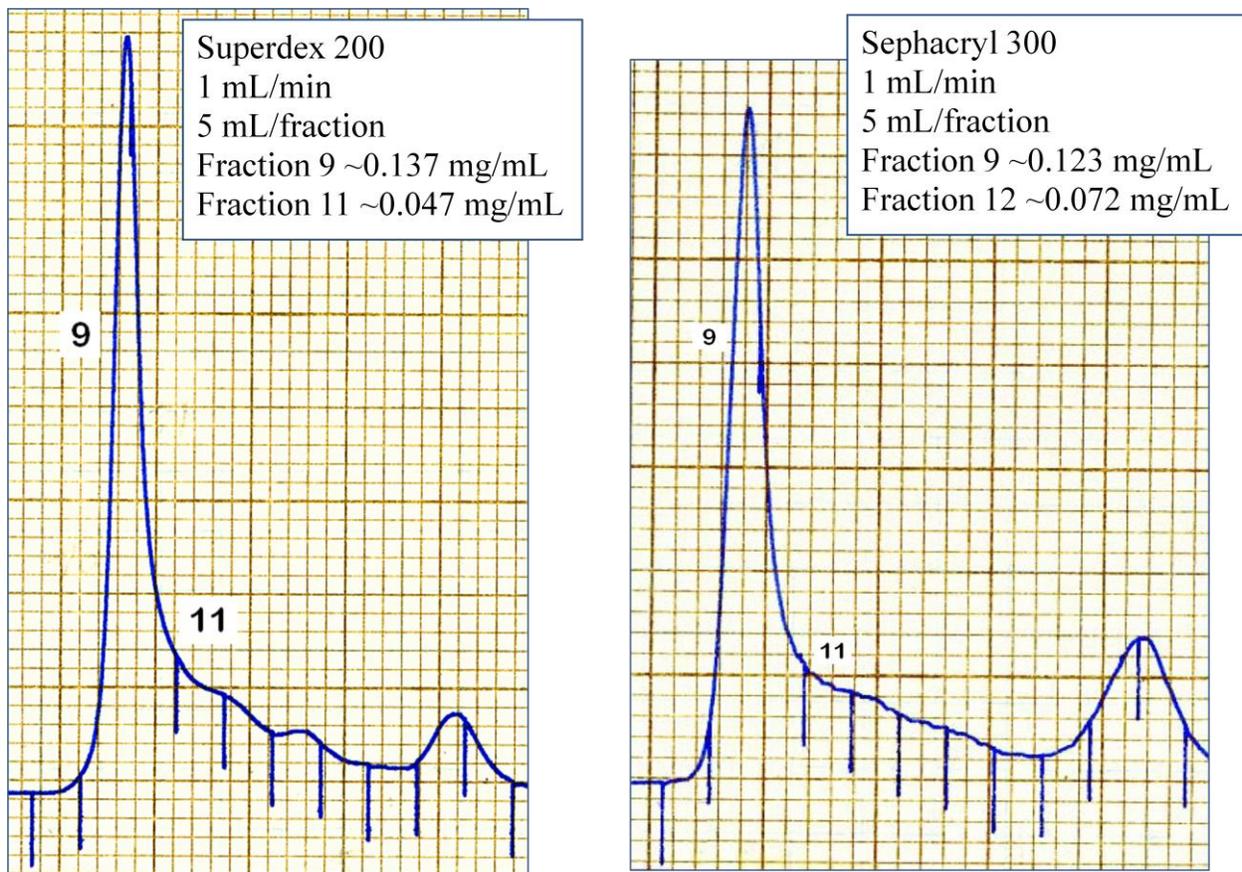


Figure III.9 Gel filtration of crude mSNAPc treated with lauryl maltoside (left) and Tween 20 (right) both showing aggregation in fraction 9 and pure protein in fraction 11.

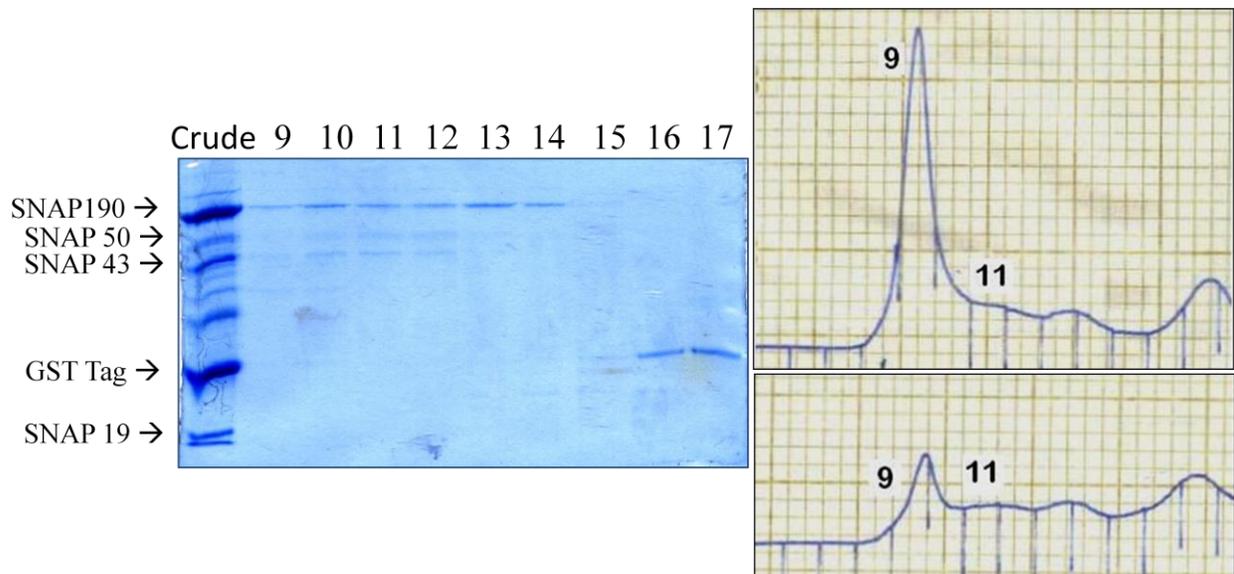


Figure III.10 Left: SDS page gel of DNase treated mSNAPc. Lane 1: Crude mSNAPc before gel filtration. Lanes 2-10: Fractions 9 – 17 of the resulting gel filtration. Right: Superdex 200 gel filtration of crude mSNAPc before (top) and after (bottom) treatment with DNase. Fraction 9 represents the void peak. Fraction 11 represents pure mSNAPc protein.

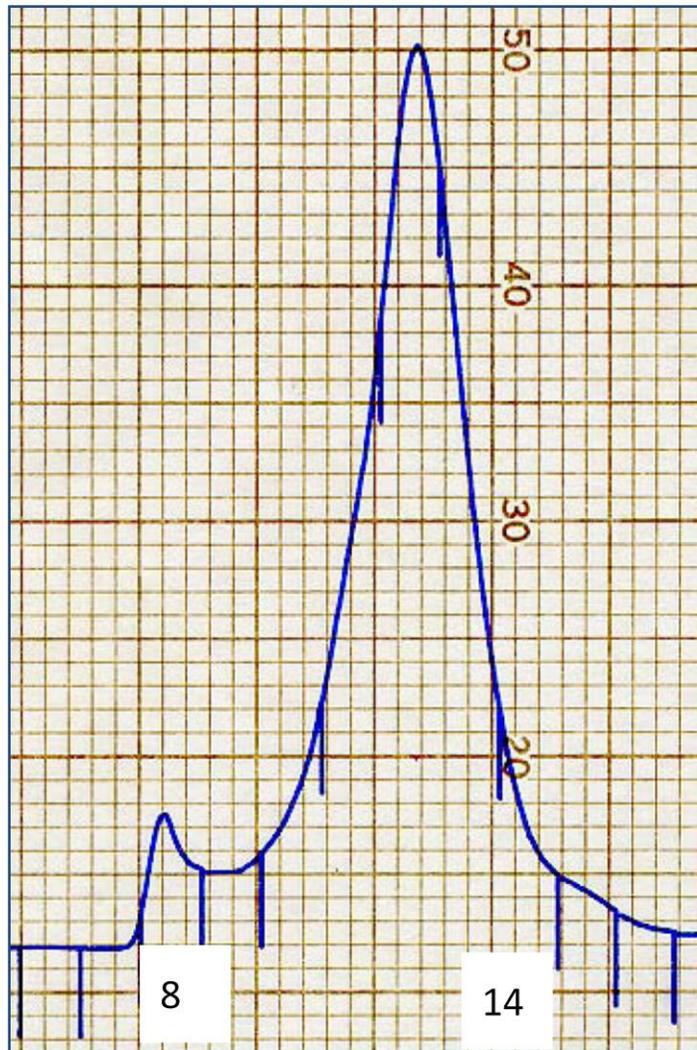


Figure III.11 Gel filtration of 4 combined previously run gel filtrations of mSNAPc representing 25 total purifications of 150 total liters of cell culture.

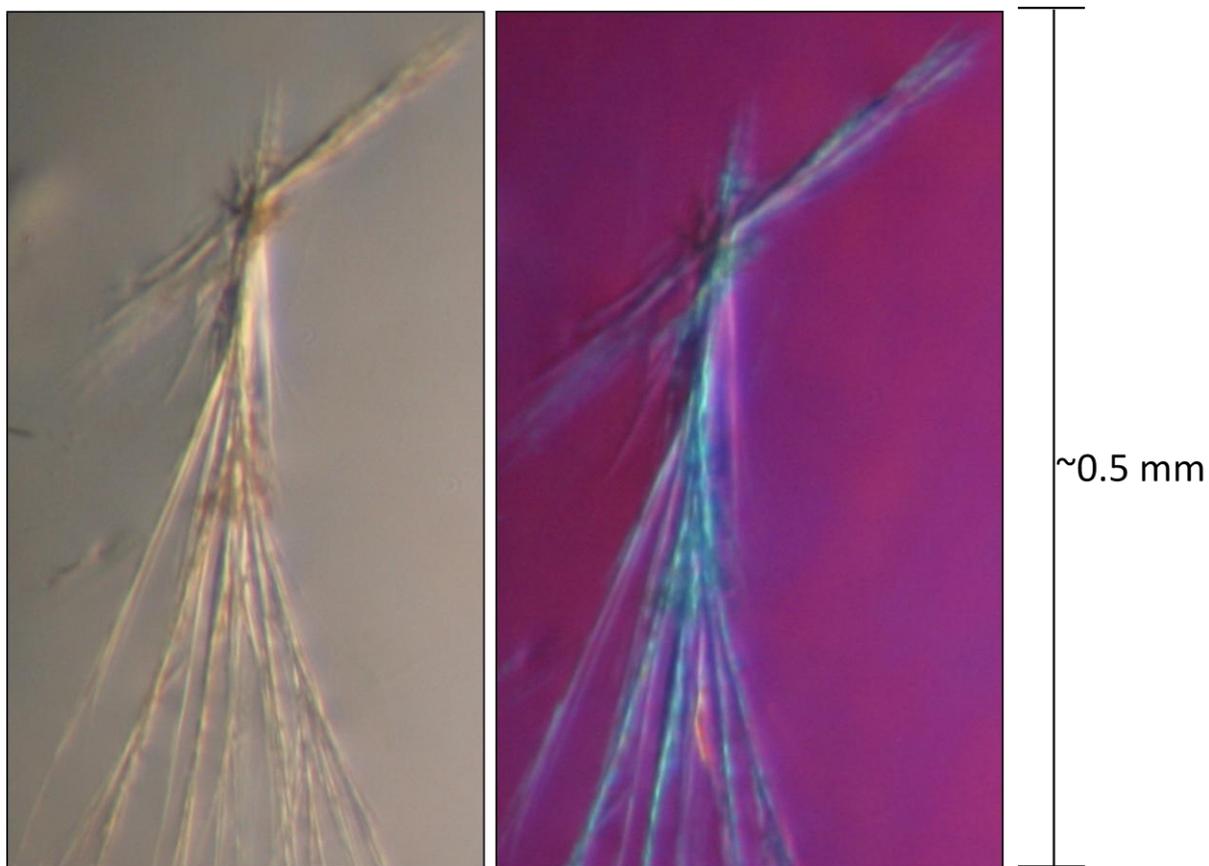


Figure III.12 Left: Needle of mSNAPc grown in 0.1 M HEPES, pH 8.0, 0.05 M $MgCl_2$, 0.2 M NaCl, and 8% PEG mme 5000. Right: The same needles as on the left under polarized light.

Purification was further optimized to potentially improve crystal quality by eliminating small contaminants. This was achieved by changing the buffer to HEG 1K (20 mM HEPES, 2 mM EDTA, 10% glycerol, 3 mM DTT, 1 M KCl) as well as adding lauryl maltoside detergent to the wash buffer. Rather than eluting the protein from the GST resin, thrombin was used to directly cleave the protein complex from the tag while it was still bound to the affinity matrix (**Figure III.13** and **Figure III.14**).

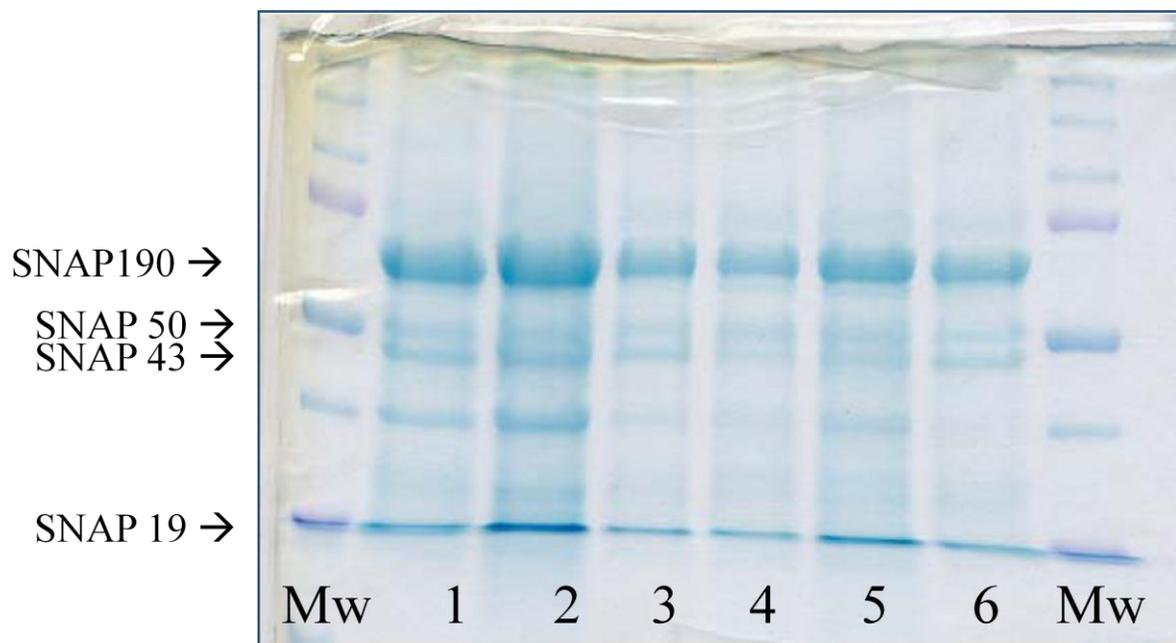


Figure III.13 SDS Page gel of crude mSNAPc. Lane Mw: Molecular weight standard. Lanes 1 and 2: Original Purification of mSNAPc. Lanes 3-5: Different redundant purifications of mSNAPc with 1M total KCl in HEG buffers. Lane 6: Purification of mSNAPc with 1M KCl total in HEG buffers plus lauryl maltoside in the wash buffer. SNAP19 is running with the salt front of the gel.

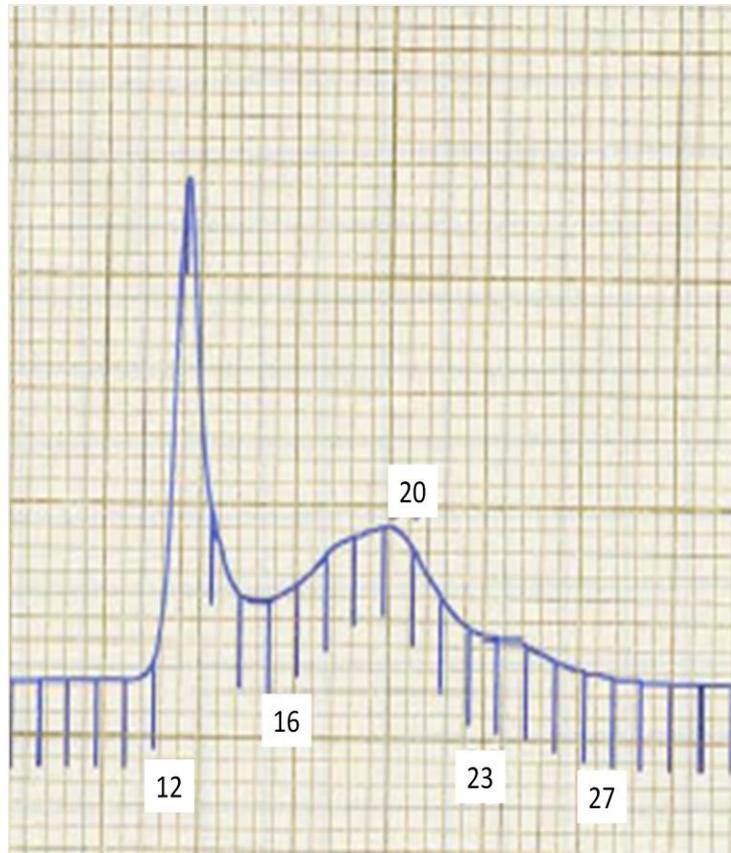


Figure III.14 Gel filtration of 6 combined purifications of crude mSNAPc purified as described above. Fraction 13 represents the void. Fraction 18 represents pure protein.

At this time additional affinity tags were used to allow for multiple sequential affinity purifications. The first tag was a non-cleavable N-terminal His-tag on the SNAP50 subunit.

From 2 liters of culture, 6.7mg of crude 50/43 were co-eluted from 5mL of Ni NTA resin

(**Figure III.15**). Crude 50/43 could not be concentrated using YM-50 Centriprep concentrators

(Millipore).

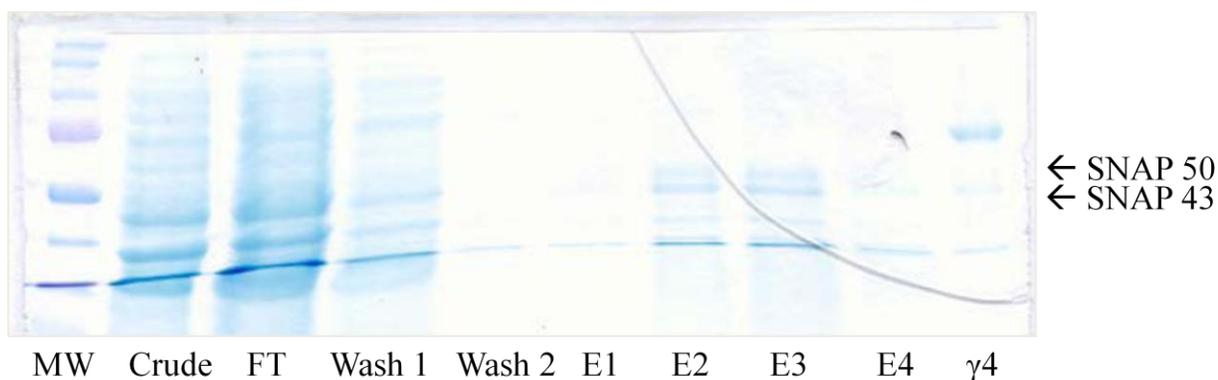


Figure III.15 SDS Page gel of co-expressed His-tagged SNAP50 and SNAP43. Lane MW: Molecular Weight. Lane Crude: Crude lysate. Lanes Wash 1 and Wash2: 20 mM imidazole containing buffer wash of Ni resin. Lanes E1-E4: 250 mM imidazole containing buffer elutions of Ni resin. Lane γ 4: Sample of pure mSNAPc showing the location of the SNAP50 and SNAP43 bands.

From 2 liters of BL-21 competent cell culture 10.9 mg of crude SNAP50/SNAP190 (1-505)/SNAP43 complex was eluted from Ni NTA resin (**Figure III.16**).

A co-expression of His-tagSNAP50, GST-tagSNAP190, SNAP45 and SNAP19 was also tested. Typical yields varied but averaged about 5 mg/L of crude protein per cell culture when Ni affinity chromatography was used. Addition of lauryl maltoside to either the lysis buffer alone or all protein contacting buffers failed to improve yields or quality of the crude protein.

A single purification of co-expression of His-tagSNAP50, GST-tagSNAP190, SNAP45 and SNAP19 yielding 2.1 mg/mL of crude complex via the His-tag, followed by further purification with GST affinity resin in HG-250 lead to increased purity of the complex with 62% recovery of the thrombin cleaved protein from the GST resin. Alternatively, expression levels of 1.2mg/mL were observed when purifying co-expressed of His-tagSNAP50, GST-tagSNAP190, SNAP45 and SNAP19 using GST affinity resin with HG-250 w/out DTT with cleavage by thrombin digestion (**Figure III.17**). Approximately 23% of this crude protein was recovered

when further purified with Ni NTA resin. This protein could be concentrated to 1.2mg/mL with 42% loss of protein

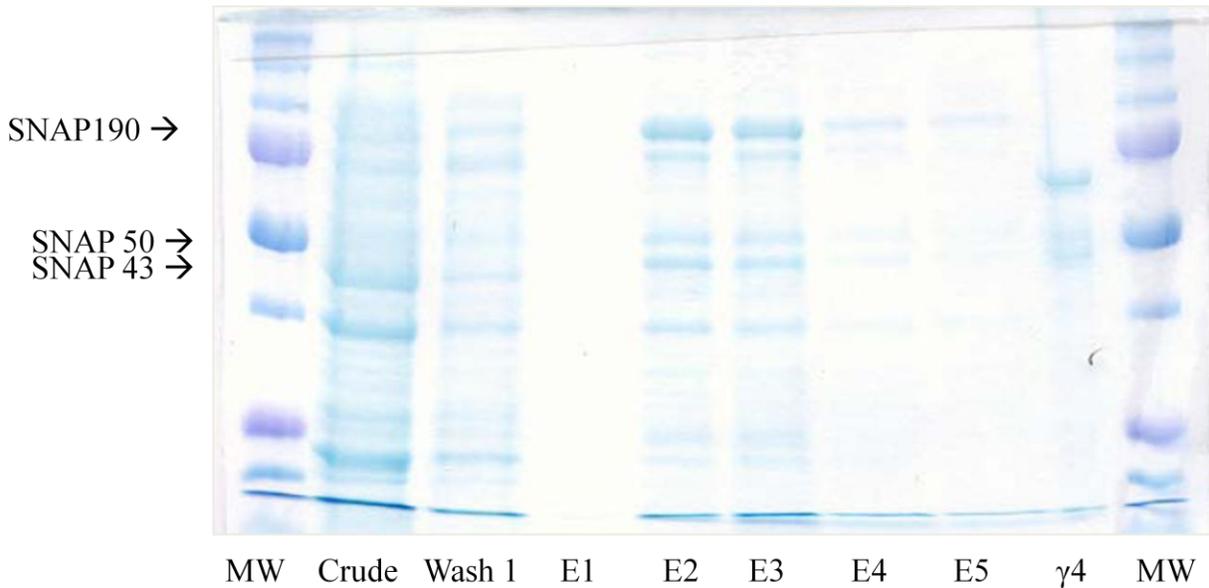


Figure III.16 SDS Page gel of co-expressed His-tagged SNAP50, SNAP43, and N-terminal GST tagged SNAP190 (1-505). Lane MW: Molecular Weight. Lane Crude: Crude lysate. Lane Wash 1: 20 mM imidazole containing buffer wash of Ni resin. Lanes E1-E4: 250 mM imidazole containing buffer elutions of Ni resin. Lane γ 4: Sample of pure mSNAPc showing the location of the SNAP50 and SNAP43 bands.

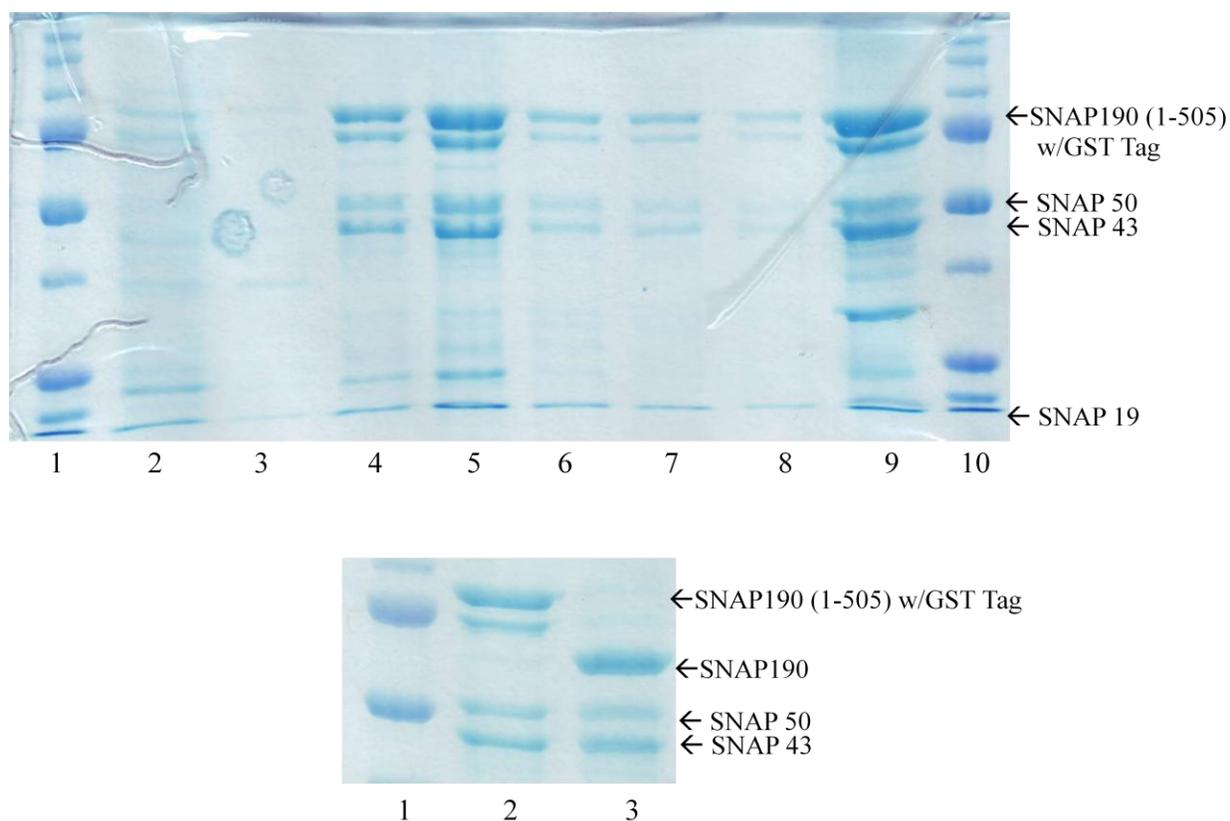


Figure III.17 Top: SDS Page gel of co-expressed SNAP50 (N-terminal His tag), SNAP43, SNAP190 (1-505 with C-terminal GST tag) and SNAP19. Lanes 1 and 10: Molecular weight standard. Lane 2: FT of wash buffer. Lane 3: FT of wash buffer plus 1% lauryl maltoside. Lanes 4-8: Ni NTA Elutions of crude protein complex. Lane 8: Ni NTA resin after elution. Bottom: SDS Page gel of co-expressed SNAP50 (N-terminal His tag), SNAP43, SNAP190 (1-505 with C-terminal GST tag) and SNAP19 further purified after Ni purification. Lane 1: Molecular weight standard. Lane 2: Crude complex that flowed through the GST resin. Lane 3: Complex that was cleaved from resin after thrombin digestion.

Since previous attempts to crystallize SNAP190 (1-505) were unsuccessful, it was theorized that the Myb domain in the absence of DNA might be too disordered for crystallization. C-terminal truncations of the SNAP190 subunit were designed to remove the Myb domain [(1-255), (1-260), (1-265)] as well as the linker connecting the Myb domain to the SNAP50 interacting domain [(1-131), (1-135)].

SNAP190 (1-131) had expression levels of 2.0mg/mL when purified as a complex with SNAP43 and SNAP50. SNAP190 (1-135) expressed at levels of 2.5 mg/mL when purified as the same complex. Expression levels of 1.0mg/mL were observed for the complex of SNAP190 (1-255). SNAP190 (1-260) expressed at 1.5 mg/mL as a SNAP50/SNAP43 complex. Crystallization trials of the individual proteins as well as the complexes were performed by lab members Rafida Nossoni and Camille Watson without success.

The GST affinity tag and thrombin linker were next removed from SNAP190 to produce a tagless SNAP190. Co-expression of the tagless SNAP190 with the C-terminal and N-terminal histidine tagged SNAP50 revealed better yields with the N-terminally tagged SNAP50. Typical yields for Ni NTA purification of the tagless SNAP190 (1-505) co-expressed with SNAP50, SNAP43 and SNAP19 were 3.6 mg/mL. Sequencing of PVDF blots could not detect full length SNAP50, SNAP43 or SNAP19. A truncated version of SNAP190 (146-505) was identified via the molecular weight of approximately 40 KDa and an N-terminal protein sequence of “MKPYFK”.

In order to increase the solubility of the SNAP190 subunit, and conceivably the entire co-expressed unit, the affinity tag was changed from GST to MBP (Maltose Binding Protein) with a Factor Xa cleavable linker. Typical yields of 11mg/mL for the protein alone or 15mg/mL of the co-expressed complex were observed, though typically the other subunits were not observed in co-expression. The Factor Xa linker was also difficult to cleave with Factor Xa.

Therefore, the linker was mutated from a Factor Xa recognition sequence to a thrombin recognition sequence. Yields of 12mg/mL crude protein or 5.3mg/mL when expressed as a complex were observed. As before, the other subunits were not observed in co-expression.

Since thrombin can digest a protein nonspecifically and the Factor Xa protease did not digest the protein properly and is expensive, the Factor Xa recognition sequence within the pMAL_C4x was changed to the Smt3 protein who's N-terminus is recognized for proteolytic degradation by the SUMO (Small Ubiquitin-like Modifier) protease. The SUMO protease, originally utilized by Christopher Lima's group, is known to cut specifically and is inexpensive to make. The Smt3 protein also helps to solubilize the proteins to which it is fused.

The pMAL_SUMO_190 fusion was purified as described above with typical yields of 10mg/L when expressed as the complex. Again, there was no evidence of co-expression of the other subunits, SNAP50, SNAP43 or SNAP19. The fusion of MBP-SMT3-SNAP190 (1-131) was generated, but never tested.

An MBP/SMT3 fusion was generated of SNAP50. This construct failed to express even with minor changes to the promoter region (**Figure III.18**).

PS_S50 Promoter Region

```
ttaccgacttcagcgagaccggttatagcctttatcgccggttaatccagattaccagttta 180
ccttcttcgattttcatcgattatctccttattaaagttaaacaaaattattttctacagg 240
ggaattggttatccgctcacaattcccctatagtgagtcgtattaatttcctaataatgcagga 300
gtcgcataagggagagcgtcgagatcccggacaccatcgaatggcgcaaaacctttcgcg 360
```

Untagged S50 Promoter Region

```
ggagactgggtcctgcctgccaccacccccgctacacgtagggccacctcggccttccttc 300
agccatggtatatactccttattaaagttaaacaaaattattttctacagggggaattggtat 360
ccgctcacaattcccctatagtgagtcgtattaatttcctaataatgcaggagtcgcataagg 420
gagagcgtcgagatcccggacaccatcgaatggcgcaaaacctttcgcggtatggcatga 480
```

PS_S50 Fixed Promoter Region

```
ttcttaccgacttcagcgagaccggttatagcctttatcgccggttaatccagattaccagt 180
ttaccttcttcgattttcatgggtatatactccttattaaagttaaacaaaattattttctac 240
aggggaattggttatccgctcacaattcccctatagtgagtcgtattaatttcctaataatgca 300
ggagtcgcataagggagagcgtcgagatcccggacaccatcgaatggcgcaaaacctttc 360
```

Figure III.18 Comparison of the pMAL_SUMO_S50/S43 (PS_S50) plasmid's promoter region (top) with that of the untagged original pCDF_S50 promoter region (middle). Differences are highlighted in green. The corrected pMAL_SUMO_S50/S43 plasmid (bottom) as determined by sequencing.

The amino acid sequence for the SNAP19 subunit appears to contain an N-terminal leucine zipper motif (1-41) as well as a high concentration of ten glutamic acid residues from amino acids 86-96 (**Figure III.19**). Stop codons were then inserted into the SNAP19 ORF to produce two truncations: SNAP19 (1-41) and SNAP19 (1-85). These truncations appeared to have no effect on the co-expression of mSNAPc.

```

78 atgctgagccggcttcaggaactgcgcaaggaggaggagacgctg
M L S R L Q E L R K E E E T L
123 ctgcggttgaaggcagccctgcacgaccagctgaaccgcctcaag
L R L K A A L H D Q L N R L K
168 gttgaagaattagccctccaatcaatgatcagttctagaagaggg
V E E L A L Q S M I S S R R G
213 gatgagatgctgtcttctcacactgtacctgaacagtcacatgat
D E M L S S H T V P E Q S H D
258 atggttggtgcatgtagacaatgaagcatcaatcaaccaaacaacc
M L V H V D N E A S I N Q T T
303 ctggagctgagcacaagagatcatgtgacggaagaggaggaggag
L E L S T K S H V T E E E E E
348 gaagaggaagaagaatcagattcctaa 374
E E E E E S D S *

```

Figure III.19 The open reading frame of SNAP19. Circles highlight the leucine zipper motif. The glutamic acid region is underlined.

Surface Entropy Reduction (SER) calculations were performed on the SNAP190 (1-505) ORF revealing possible sites for mutation. The triple mutant K294A/Q295A/E296A was selected and mutation achieved via mutagenesis. Despite efforts to express and purify this construct under various conditions, there was never any evidence of protein production.

A MBP-SMT3-SNAP190 (1-131) and SNAP190 (260-505) fusion with various lengths of poly-glycine/serine linking the two was next created (**Figure III.20**). These fusions would remove a segment predicted to be disordered while keeping segments known to have important interactions between the subunits (SNAP190 and SNAP50) and DNA (the MYB domain). These fusions failed to express and were ultimately handed over to Dr. Stacy Hovde for further testing.

The final construct designed was a Maltose Binding Protein tagged SNAP190 (Δ 131-260) with a SMT3 cut site. Completion of this construct and testing were completed by Dr. Stacy Hovde.

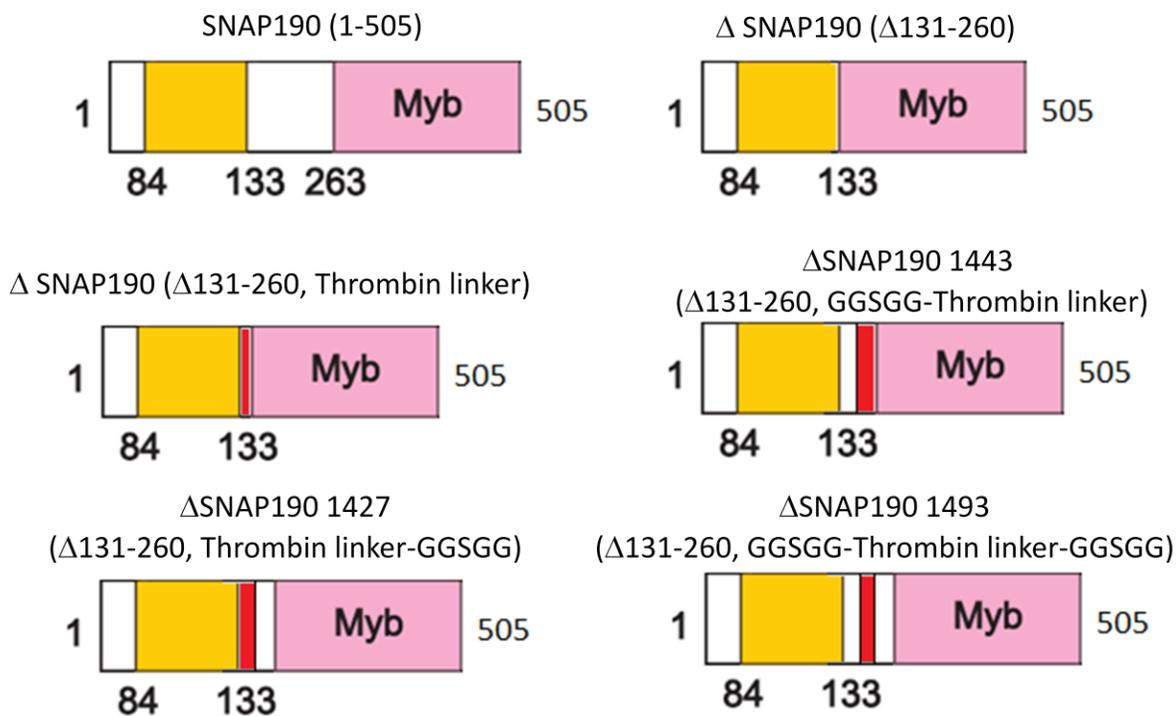


Figure III.20 Cartoon representations of the different delta constructs. Pink: Myb domain. Yellow: SNAP190-SNAP50 interacting region. Red: Thrombin cleavable linker.

In total, several parameters were explored including different cell lines for expression (RP, RIL and Rosetta), buffers, resins, growing conditions, truncations and affinity tag combinations. Of these that expressed, none of these were able to improve the stoichiometry between the four subunits of mSNAPc or completely eliminate the aggregation issue. Therefore, crystals suitable for x-ray diffraction are still elusive.

Some possible solutions are still being pursued, though a more exhaustive search of expression systems and vectors might prove valuable. Such services are currently being developed by such groups as Dr. W. C. Brown's at the University of Michigan. With more optimization at the expression level it might still be possible to obtain enough pure, stoichiometrically pure protein.

REFERENCES

REFERENCES

- 1 Voet, Donald, and Judith G. Voet, Biochemistry, Third ed. Hoboken, NJ: John Wiley & Sons, Inc., 2004.
- 2 Cooper TA (2009). "RNA and disease". *Cell*. **136** (4): 777-793.
- 3 Myer VE, Young RA (1998). "RNA polymerase II holoenzymes and subcomplexes". *J. Biol. Chem.* **273** (43): 27757–60.
- 4 Kornberg R (1999). "Eukaryotic transcriptional control". *Trends in Cell Biology*. **9** (12): M46.
- 5 Sims RJ 3rd, Mandal SS, Reinberg D (2004). "Recent highlights of RNA-polymerase-II-mediated transcription". *Current opinion in cell biology*. **16** (3): 263–271.
- 6 Meyer PA, Ye P, Zhang M, Suh MH, Fu J (2009). Structure of the 12-subunit RNA polymerase II refined with the aid of anomalous diffraction data". *J.Biol.Chem.* **284**: 12933–12939.
- 7 Schramm L, Hernandez N (2002) "Recruitment of RNA polymerase III to its target promoters". *Genes & Development*. **16**: 2593-2620.
- 8 Geiduschek EP, Kassavetis GA (2001). "The RNA polymerase III transcription apparatus". *J. Mol. Biol.* **310**: 1-26.
- 9 Hernandez N (2001). "Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription". *J. Biol. Chem.* **276**: 26733-26736.
- 10 Henry RW, Ford E, Mital R, Mittal V, Hernandez N (1998). "Crossing the line between RNA polymerases: transcription of human snRNA genes by RNA polymerases II and III". *Cold Spring Harbor Symposia on Quantitative Biology: Mechanisms of Transcription*. **63**: 111-120.
- 11 Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A (2007). "The expanding RNA polymerase III transcriptome". *Trends in Genetics*. **23** (12): 614-622.
- 12 Riethoven JM (2010). "Regulatory regions in DNA: promoters, enhancers, silencers, and insulators". *Methods in Molecular Biology: Computational Biology of Transcription Factor Binding*. **674**: 33-42.
- 13 Kuhlman TC, Cho H, Reinberg D, Hernandez N (1999). "The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter". *Molecular and Cellular Biology*. **19** (3): 2130-2141.

- 14 Ma B, Hernandez N (2001). "A map of protein-protein contacts within the small nuclear RNA-activating protein complex SNAPc". *Journal of Biological Chemistry*. **276** (7): 5027-5035.
- 15 Schuster C, Myslinski E, Krol A, Carbon P (1995). "Staf, a novel zinc finger protein that activates the RNA polymerase III promoter of the selenocysteine tRNA gene". *EMBO*. **14** (15): 3777-3787.
- 16 Conaway JW, Conaway RC (2004). "RNA polymerase II and basal transcription factors in eukaryotes". *Encyclopedia of Biological Chemistry*. **3**: 763-765.
- 17 Lobo SM, Lister J, Sullivan ML, Hernandez N (1991), "The cloned RNA polymerase II transcription factor IID selects RNA polymerase III to transcribe the human U6 gene in vitro". *Genes & Development*. **5** (8): 1477-1489.
- 18 Henry RW, Kobayashi R (1995). "SNAPc. A novel TBP-TAF complex required for transcription by RNA polymerases II and III". *Experimental Medicine*. **13** (18): 2179-2182.
- 19 Lee TI, Young RA (1998). "Regulation of gene expression by TBP-associated proteins". *Genes & Development*. **12** (10): 1398-1408.
- 20 Geiduschek EP, Kassavetis GA (1995). "Comparing transcriptional initiation by RNA polymerases I and III". *Current Opinion in Cell Biology*. **7** (3): 344-351.
- 21 Henry RW, Ma B, Sadowski CL, Kobayashi R, Hernandez N (1996) "Cloning and characterization of SNAP50, a subunit of the snRNA-activating protein complex SNAPc". *EMBO J*. **15**: 7129-7136.
- 22 Jawdekar GW, Hanzlowsky A, Hovde S L, Jelencic B, Feig M, Geiger J H, Henry RW (2006) "The unorthodox SNAP50 zinc finger domain contributes to cooperative promoter recognition by human SNAPc". *J. Bio. Chem*. **281** (41): 31050–31060.
- 23 Hinkley CS, Hirsch HA, Gu L, LaMere B, Henry RW (2003). "The small nuclear RNA-activating protein 190 Myb DNA-binding domain stimulates TATA box-binding protein-TATA box recognition". *J. Bio. Chem*. **278** (20): 18649-18657.
- 24 Hanzlowsky A, Jelencic B, Jawdekar G, Hinkley CS, Geiger JH, Henry RW (2006) "Co-expression of multiple subunits enables recombinant SNAPc assembly and function for transcription by human RNA polymerases II and III". *Protein Expression and Purification*. **48**: 215–223.

Chapter IV: TFIIB Brf1-TBP Triple Fusions

IV.1. Background

IV.1.1. TFIIB

Transcription Factor IIB (TFIIB) is a central transcription initiation factor of RNA polymerase III (RNAPIII) (1-3) (see Section III.1.1.). It recruits RNAPIII to the transcription initiation start site as part of several different pre-initiation complexes (PICs) depending on the arrangement of the promoter elements to which RNAPIII is being recruited. The 5S rRNA promoter represents a type 1 promoter and contains a gene internal box C element (also known as an internal control region) which is recognized by TFIIC which in turn recruits TFIIB followed by Brf1-TFIIB. Box A and box B elements within the tRNA ORF represent type 2 gene internal sequences recognized by TFIIC which directly recruits TFIIB. TFIIB can also be recruited to the promoter via its TBP subunit if a gene external TATA box element is introduced, as is the case with the U6 snRNA promoter. In all cases RNAPIII is recruited after the introduction of TFIIB (4, 5). It is not well understood how TFIIB is able to distinguish between the different promoter arrangements in order to properly recruit RNAPIII.

TFIIB is composed of three subunits: TATA Binding Protein (TBP), TFIIB Related Factor 1 (Brf1), B'' 1 (B Double Prime 1, Bdp1). All three are known to interact strongly with one another even in the absence of the promoter, though the TBP and BRF1 interaction is the tightest with the two often co-purifying as a complex known as B' in *Saccharomyces cerevisiae* (6). It is TBP that interacts with the TATA box of the promoter elements. Brf1 has a distinct TFIIB-related N-terminal domain comprising two cyclin fold repeats (aa 94-164 and 189 - 264) as well as three Brf homology regions (aa 282 – 596) on the C-terminus (7).

Since the recruitment of RNAP III depends on the recruitment of TFIIB for Pol III-specific genes, determining the structure of TFIIB would aid in understanding how TFIIB interacts with the other units of the PICs, the promoter and ultimately RNAP III.

IV.1.2. Creation of the BRF1-TBP Triple Fusion

Brf1 in budding yeast binds TBP through an N-terminal domain (aa 1-282 or 1-365) as well as a C-terminal domain (aa 439-545). In addition, two particular yeast BRF homologues, *Candida albicans* and *Kluyveromyces*, lack aa 366-407 as well as aa 383-424 respectively (7). Though the entire structure of TFIIB has yet to be determined, several segments have been determined to atomic resolution. Particular to our investigation was that of a human TFIIBc-TBPc complex bound to an idealized and extended promoter (8) (**Figure VI.1**). The second of note is that of yeast Brf1-TBP-DNA ternary complex (9). It includes the TBPc (aa 61-240) and the c-terminus of Brfc (aa 439-596). This combined with the known structure fragments of TFIIB allowed for several fusions of the two to be generated by our collaborators (6).

The fusions consist of the TBP associating N-terminal domain of BRF1 fused to the C-terminal domain of the core element of TBP (aa 62-240) (10) which in turn is fused to the TBP associating C-terminal domain of BRF1 (**Figure IV.2**). These fusions are able to bind Bdp1 and successfully recruit RNAPIII to the promoter *in vivo* (6). In addition to their activity, they represent a desirable target for crystallizations due to the fact that the different units are linked together stoichiometrically.

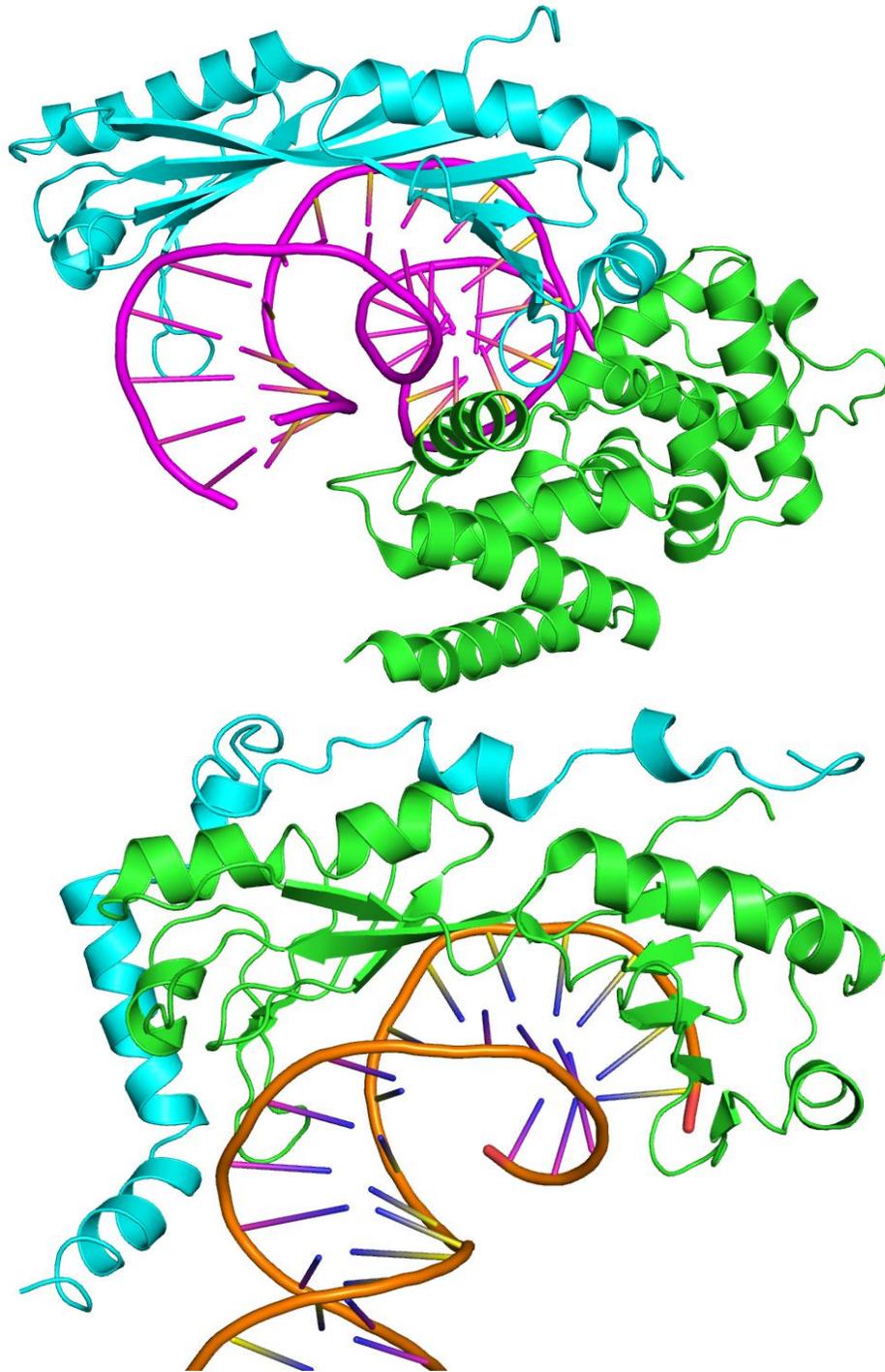


Figure IV.1 Top: Crystal structure of a human TBP core domain (blue)-human TFIIB core domain (green) complex bound to an extended, modified adenoviral major late promoter (orange) (8) (PDB ID 1C9B). Bottom: a yeast Brf1 (blue)-TBP (green)-DNA (orange) ternary complex (9) (PDB ID 1NGM). Aligning the TBP core domains gives the relative orientation of the other segments when designing fusion constructs.

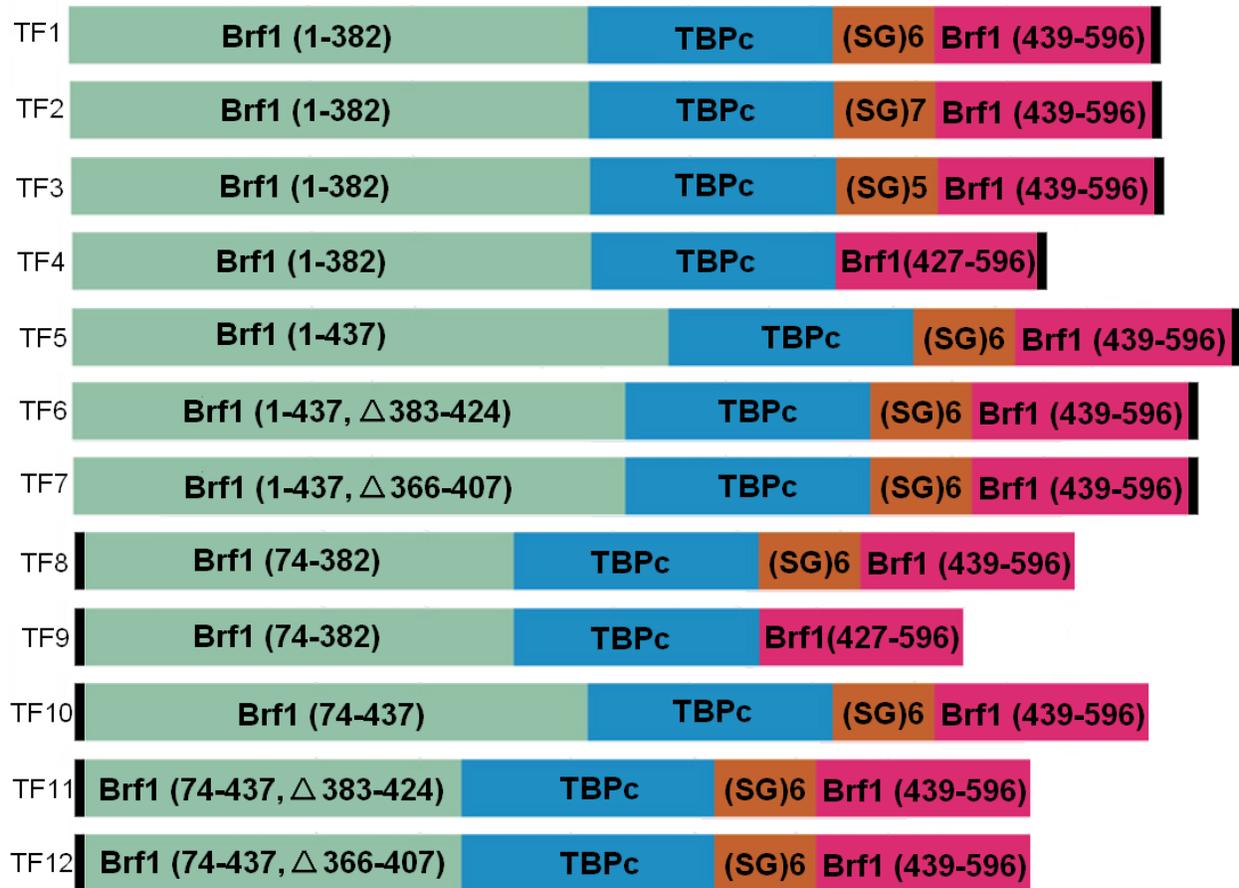


Figure IV.2 Cartoon representation of the ORFs for several different TF fusion constructs. The black bars represent the locations of poly-histidine affinity tags.

IV.2. Experimental Procedures

IV.2.1. Growth Optimization

Growth was adapted from a protocol from the Kassavetis lab. The plasmids containing the ORF's of TF1 and TF8 were first transformed into competent Bl21-CodonPlus (DE3)-RIPL cells. Single colonies were selected from ampicillin containing agar plates and allowed to grow over night to saturation. To these broths 100% glycerol was added to bring the total concentration to 20%. One milliliter aliquots were flash frozen and stored at -80°C.

Six 250mL Erlenmeyer flasks containing 50mL each of LB or TB were supplemented with 50mg/mL ampicillin and 50mg/mL chloramphenicol. To these 150µL of glycerol stock was added and the flasks were allowed to shake ON at 37°C. The next morning each flask was individually transferred to 1L of TB containing 50mg/mL ampicillin and 50mg/mL chloramphenicol. The flasks shook at 37°C until an OD=0.9 was reached. The cells were then cooled to 16°C using a refrigerated shaker and induced with 0.1 mM IPTG.

Following overnight expression, the cells were harvested by centrifugation at 5K RPM for 20 minutes. The resulting cell pellets were either used fresh or frozen at -20°C until use.

IV.2.2. Purification Optimization

Purification was adapted from a protocol from the Kassavetis lab. First, 6L worth of frozen cell pellets were suspended in 200mL of Lysis Buffer (50 mM HEPES, pH = 7.8, 1.14 M NaCl, 20 mM imidazole, 5% glycerol, 0.5 mM PMSF and 10 mM β-mercaptoethanol) supplemented with 4 Complete protease inhibitor tablet (Roche). The cells were lysed on ice using a Branson Sonifier for three 60 second cycles with 60 seconds of rest in between. The lysate was then clarified at 15K RPM for 30 minutes.

The clarified lysate was then bound to 15 mL Ni NTA resin. The resin was washed with Wash Buffer (20 mM HEPES, pH = 7.8, 7 mM MgCl₂, 500 mM NaCl, 20 mM imidazole, 10% glycerol, 10 mM β-mercaptoethanol and 0.5 mM PMSF) until a protein concentration less than 0.1 was detected in the flow through. Protein was then eluted using Elution Buffer (same as Wash Buffer but with 200 mM imidazole).

The protein was then buffer exchanged using dialysis membrane into Buffer H plus 300 mM NaCl (40 mM HEPES (pH 7.8), 0.2 mM EDTA, 10% glycerol, 10 mM 2-mercaptoethanol, 0.5 mM PMSF) for 4 hours. The dialyzed protein was then loaded into a 300 mM NaCl Buffer H equilibrated HiTrapHeparin HP column at a rate of 1 mL/min.

The column was washed with 5 mL of Buffer H containing 300 mM NaCl followed by a 10mL gradient from Buffer H plus 300 mM NaCl to Buffer H plus 400 mM NaCl. The column is developed using a gradient from 400 to 800 mM NaCl of 40 mL of Buffer H. The TF fusions typically elute between 600 and 650 mM NaCl. As a final step the column is washed with 10 M NaCl containing Buffer H.

Fractions containing TF fusion as determined by SDS Page analysis were collected, pooled and concentrated to approximately 3.2 mg/mL. The appropriate amount of concentrated DNA was added to achieve a 1:1.2 protein: DNA ratio and left to sit for 30 minutes on ice. The mixture was then concentrated using a Vivaspin 500 50 KDa MWCO concentrator. This sample was then applied to a BioRad RNase-Free Micro Bio-Spin Column P-30 equilibrated with the final buffer, B500-75 (10 mM Tris, pH 8.0, 75 mM NaCl, 15% glycerol, 5 mM DTT). The recovered protein is again concentrated using the Vivaspin concentrators to no higher than 5 mg/mL.

DNA used was ordered from the Macromolecular Structure Facility at Michigan State University. All strands were HPLC purified against varying sodium chloride salt gradients customized to each strand. Fractions containing the DNA were diluted to low salt and concentrated using DiEthylAminoEthane (DEAE) resin. Further concentration was achieved using Centricon NMWL 10,000 concentrators (Millipore). Final concentration was determined by UV absorption at 260 nm. Equal amounts of complimentary strands were combined and then annealed by placed the vessel in boiling water which was then allowed to cool to room temperature.

IV.3. Results and Discussion

IV.3.1. Purification and attempted crystallization of TBP-Brf1 fusions

First the growth of the cell cultures was adapted from a protocol from the Kassavetis lab. Several broths including LB (Luria Broth), TB (Terrific Broth) and BM (Base Media, described previously) were considered for growth. Tests revealed the greatest yields of protein (almost 3.0mg/mL compared to 1.2mg/mL for LB and 0.6mg/mL for BM) came from TB (13.3g Tryptone, 26.7g Yeast extract, 4.5mL glycerol, 100mL/L of 0.18M potassium phosphate monobasic and 0.79M potassium phosphate dibasic autoclaved separately).

Next, purification of the proteins (TF1 and TF8) was adapted from a protocol from the Kassavetis lab. All fusions contained a poly-His affinity tag. Typical yields of crude TF1 and TF8 from Ni NTA affinity chromatography were approximately 13mg/L (**Figure IV.3**).

Both TF1 and TF8 were originally tested, but TF1 was difficult to purify. Therefore the focus turned to TF8, which had been previously reported to produce crystals. None of the other ten TF constructs were tested and represent an untapped potential for further study.

Since TF8 harbors a His tag, purification was initiated by affinity Ni-NTA chromatography. Imidazole concentration in the wash steps higher than 20 mM was detrimental to this step. High clarification speeds of 15,000 RPM tended to give cleaner crude Ni NTA elutions. It is also worth noting that the fusions by their nature are highly sensitive to proteolysis, so care must be made to add appropriate protease inhibitors during the initial lysis steps.

After the initial Ni NTA purification, the protein was estimated to be 80% pure as evident by SDS-Page (**Figure IV.3**). The crude protein would then be dialyzed into low salt buffer. Reduction of the salt concentration for the second step of purification, ion exchange chromatography, was particularly problematic due to precipitation of the protein in low salt, typically anything lower than 300 mM NaCl. Dilution with zero or very low salt buffers immediately precipitated the protein where the buffer was added. Dialysis into 300 mM NaCl containing buffer for longer than 4 hours caused precipitation of 50-100% of the crude protein. The use of dialysis membrane provided the best result in prevented protein precipitation in the low salt buffer.

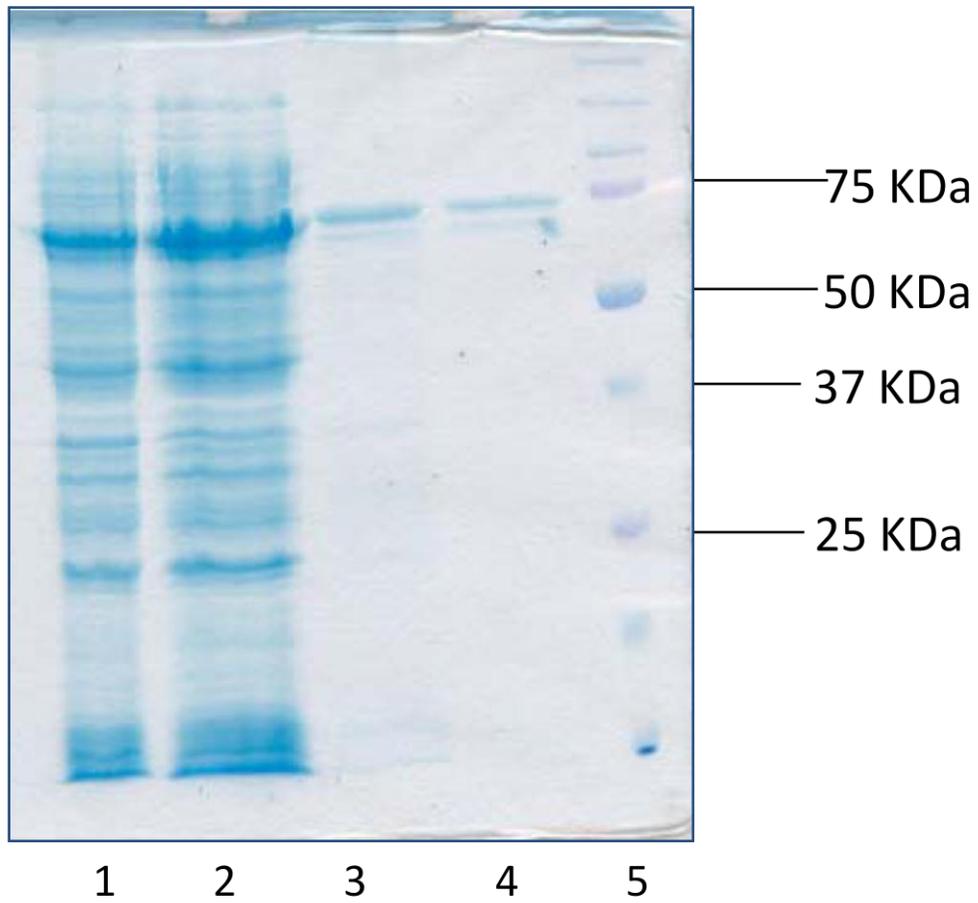


Figure IV.3 SDS Page gel of a typical TF8 purification. Lane 1: Lysate pellet. Lane 2: Lysate supernatant. Lane 3: Ni NTA resin with protein bound. Lane 4: Elution of TF8. Lane 5: Molecular weight standard.

The dialyzed protein was then loaded into a HiTrapHeparin HP column which produced two pools of protein (**Figure IV.4**). The larger fraction consisted of protein which flowed through the column irreversibly. The second fraction consisted of the minority (approximately 10%) of the protein which eluded off the HiTrapHeparin HP column at approximately 650 mM NaCl salt concentration. The flow-through fractions were tested for crystallization, but did not produce crystals. Only TF fusion found to bind to the column produced crystals. Though onerous it was possible to purify TF8 in low yield (1mg from 1 L of cell culture).

Next, it is necessary to concentrate the protein for crystallization and DNA binding. Initially, Millipore Centricon YM-50 concentrators were used after the Heparin column to concentrate the protein, but these tended to run very slow so that concentration of the pure Heparin fractions could take several days. Ammonium sulfate precipitation as a means of concentration to low salt was also tried, but much of the precipitated protein would not return to solution. The switch to Vivaspin concentrators significantly sped up this time and resulted in little loss of protein.

In order to lower the salt concentration for DNA binding, as it is known that protein/DNA complexes are generally sensitive to high salt, BioRad RNase-Free Micro Bio-Spin Columns worked the best. Dialysis was not appropriate for the size of the samples, and dilution with the lower salt buffer tended to precipitate the protein. Apparent and significant protein loss during this step was indicated by protein concentration readings using the Bradford method. However, the samples could be concentrated and protein concentration readings as low as 0.1mg/mL produced crystals. Higher protein concentrations are recommended.

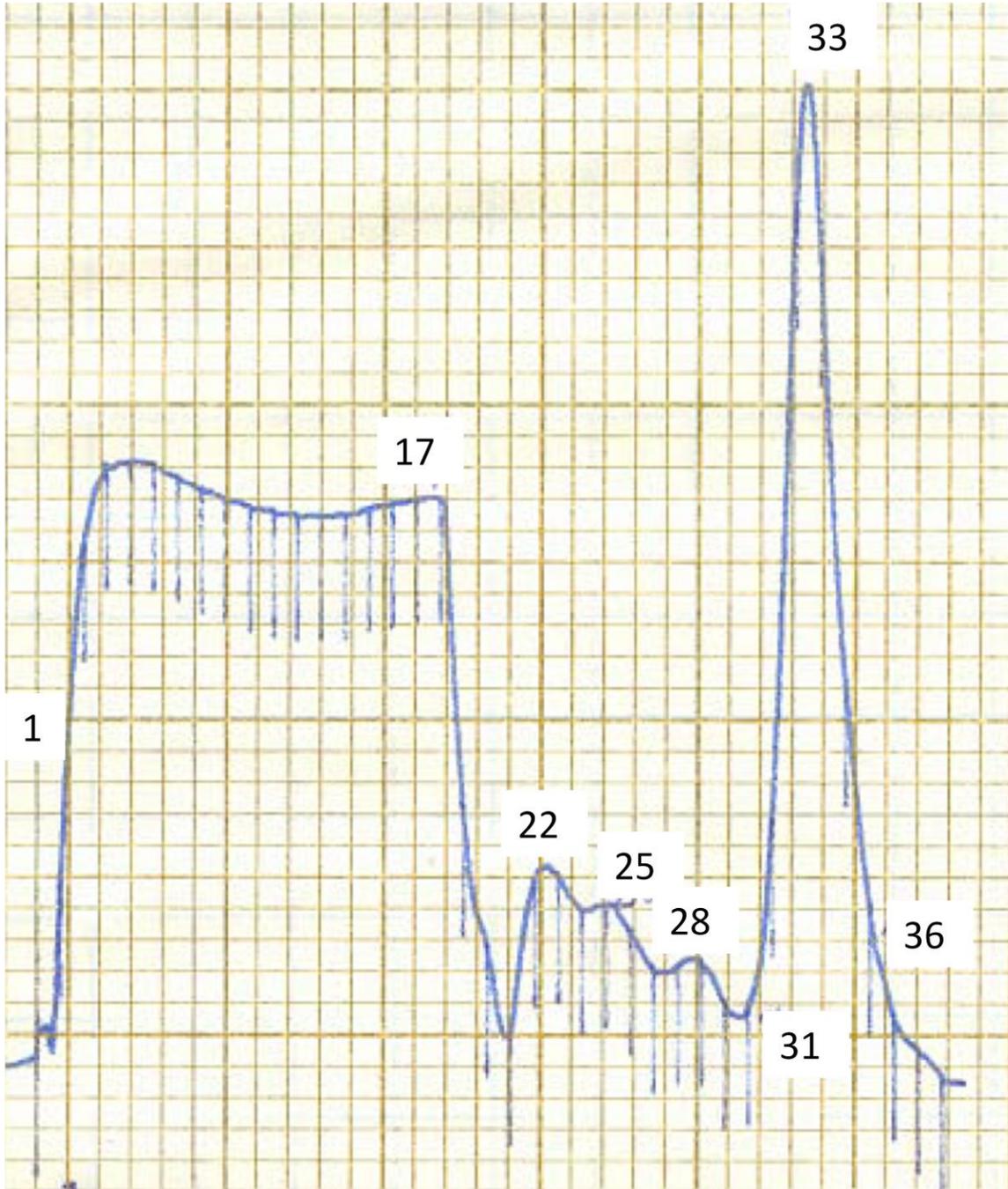


Figure IV.4 Chromatogram of HiTrapHeparin HP of TF8. Fractions 1-17 represent the load. Fractions 21-36 represent the salt gradient from 400-800 mM NaCl. Fraction 33 contains the major fraction of pure TF8.

IV.3.2. Crystallization of TF8-DNA complexes

In addition to SymSelex and SymSelex2 sequences given to us by the Kassavetis Lab, the following palindromic, nicked DNA strands were used based on the SymSelex and SymSelex2 variants (**Table IV.1**). They contain a central TATA repeat to which TBP interacts. Longer DNA pairs with large overhangs tended to cause less protein precipitation than the shorter, blunt DNA pairings for TF8.

TF1 and TF8 fusions were screened with the known crystallization conditions provided by the Kassavetis Lab (50 μ M TF:DNA complex in 10 mM Tris HCl (pH 8.0), 75 mM NaCl, 2 mM DTT, 1% glycerol with 80 mM Tris HCl (pH 8.0), 10% PEG 4K, 2 mM spermine, +/- 75 μ M ZnCl₂ reservoir buffer) using the hanging drop diffusion method. SymSelex and SymSelex2 were initially screened. The crystals were produced at room temperature after two weeks but were not suitable for x-ray diffraction (**Figure IV.5**). The screens were repeated at 4°C, but did not produce crystals. Additionally 96 well plate sitting drop screens of TF8 with the SymSelex variants were tested using all available matrix screens but these also failed to produce crystals.

Additional screening of longer, double over hang nicked SymSelex variants might produce crystals of higher quality. There also remains the possibility that one of the other TF fusion constructs not yet tested would produce protein in higher yields or that is less sensitive to low salt conditions.

Table IV.1 DNA strands derived from SymSelex and SymSelex2 sequences. The lowercase “p” represents a site of phosphorylation.

GAACGGGGTpATATAT-ACCCCGTTC SymSelex
CTTGCCCA-TATATApTGGGGCAAG 24mer

GAAACGGAGGTpATATATAT-ACCTCCGTTTC SymSelex2
CTTTGCCTCCA-TATATATApTGGAGGCAAAG 30mer

CGGAGGTpATATATAT-ACCTCCG SymSelex2-22b
GCCTCCA-TATATATApTGGAGGC 22mer

ACGGAGGTpATATATAT-ACCTCCGT SymSelex2-24b
TGCCTCCA-TATATATApTGGAGGCA 24mer

CGGAGGTpATATATAT-ACCTCCGT SymSelex2-23s
TGCCTCCA-TATATATApTGGAGGC 23mer

GGAGGTpATATATAT-ACCTCCGTA SymSelex2-24s
ATGCCTCCA-TATATATApTGGAGG 24mer

AACGGAGGTpATATATAT-ACCTCCGTT SymSelex2-26b
TTGCCTCCA-TATATATApTGGAGGCAA

ACGGAGGTpATATATAT-ACCTCCGTT SymSelex2-25s
TTGCCTCCA-TATATATApTGGAGGCA

CGGAGGTpATATATAT-ACCTCCGTA SymSelex2-24s
ATGCCTCCA-TATATATApTGGAGGC

AAACGGAGGTpATATATAT-ACCTCCGTTT SymSelex2-28b
TTTGCCTCCA-TATATATApTGGAGGCAA

AACGGAGGTpATATATAT-ACCTCCGTTT SymSelex2-27s
TTTGCCTCCA-TATATATApTGGAGGCAA

ACGGAGGTpATATATAT-ACCTCCGTTA SymSelex2-26s
ATTGCCTCCA-TATATATApTGGAGGCA

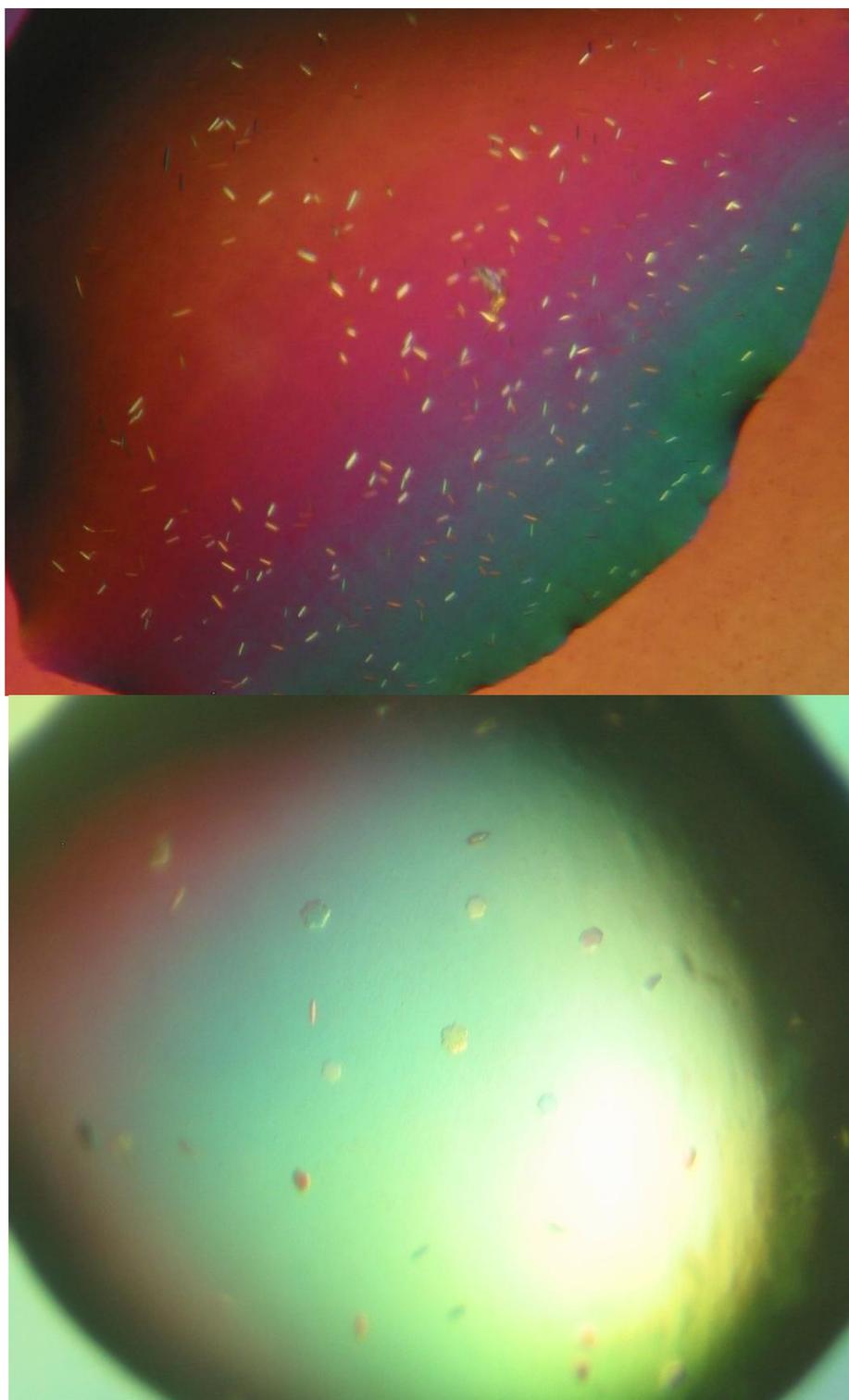


Figure IV.5 Top: Crystals of TF1 annealed to SymSelex2. Bottom: Crystals of TF8 annealed to SymSelex2.

REFERENCES

REFERENCES

- 1 Kassavetis GA, Geiduschek EP (2006). "Transcription factor TFIIB and transcription by RNA polymerase III". *Biochemical Society Transactions*. **34** (6): 1082-1087.
- 2 White RJ (2004). "RNA polymerase III transcription and cancer". *Oncogene*. **23** (18): 3208-3216.
- 3 Johnson SAS, Dubeau, L, White RJ, Johnson DL (2003). "The TATA-binding protein as a regulator of cellular transformation". *Cell Cycle*. **2** (5): 442-444.
- 4 Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A (2007). "The expanding RNA polymerase III transcriptome". *Trends Genet*. **23** (12): 614–22.
- 5 Kassavetis GA, Geiduschek EP (2006) "Transcription factor TFIIB and transcription by RNA polymerase III". *Biochemical Society Transactions*. **34** (6): 1086-1091.
- 6 Kassavetis GA, Soragni E, Driscoll R, Geiduschek EP (2005). "Reconfiguring the connectivity of a multiprotein complex: Fusion of yeast TATA-binding protein with Brf1, and the function of transcription factor IIB". *PNAS*. **102** (43): 15406-15411.
- 7 Schroder O, O.Bryant G, Geiduschek EP, Berk A, Kassavetis GA (2003). "A common site on TBP for transcription by RNA polymerases II and III". *EMBO Journal*. **22** (19): 5115-5124.
- 8 Tsai TFF, Sigler PB (2000). "Structural basis of preinitiation complex assembly on human Pol II promoters". *EMBO Journal*. **19** (1): 25–36.
- 9 Sean JZ, Kassavetis GA, Wang J, Geiduschek EP, Sigler P (2003). "Crystal structure of a transcription factor IIB core interface ternary complex". *Nature*. **422**: 534–539.
- 10 Kim Y, Geiger JH, Hahn S, Sigler P (1993). "Crystal structure of a yeast TBP/TATA-box complex". *Nature*. **365**: 512-365.